

## Abstract

Title of Document:                   MANIPULATION OF DNA TOPOLOGY  
  USING AN ARTIFICIAL DNA-LOOPING  
  PROTEIN

Daniel B. Gowetski, Ph.D., 2012

Directed By:                         Associate Professor Jason Kahn, Department of  
  Chemistry and Biochemistry

DNA loop formation, mediated by protein binding, plays a broad range of roles in cellular function from gene regulation to genome compaction. While DNA flexibility has been well investigated, there has been controversy in assessing the flexibility of very small loops. We have engineered a pair of artificial coiled-coil DNA looping proteins (LZD73 and LZD87), with minimal inherent flexibility, to better understand the nature of DNA behavior in loops of less than 460 bp. Ring closure experiments (DNA cyclization) were used to observe induced topological changes in DNA upon binding to and looping around the engineered proteins. The length of DNA required to form a loop in our artificially rigid system was found to be substantially longer than loops formed with natural proteins *in vivo*. This suggests the inherent flexibility of natural looping proteins plays a substantial role in stabilizing small loop formation. Additionally, by incrementally varying the binding site

separation between 435 bp and 458 bp, it was observed that the LZD proteins could predictably manipulate the DNA topology. At the lengths evaluated, the distribution of topological products correlates to the helical repeat of the double helix (10.5 bp). The dependence on binding site periodicity is an unequivocal demonstration of DNA looping and represents the first application of a rigid artificial protein in this capacity. By constructing these DNA looping proteins, we have created a platform for addressing DNA flexibility in regards to DNA looping. Future applications for this technology include a vigorous study of the lower limits of DNA length during loop formation and the use of these proteins in assembling protein:DNA nanostructures.

MANIPULATION OF DNA TOPOLOGY USING AN ARTIFICIAL  
DNA-LOOPING PROTEIN

By

Daniel Bernard Gowetski

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2012

Advisory Committee:

Associate Professor Jason Kahn, Chair/Advisor

Professor Jeffrey Davis

Associate Professor Douglas Julin

Distinguished University Professor George Lorimer

Professor Steven Hutcheson, Dean's Representative

© Copyright by

Daniel B. Gowetski

2012



## Dedication

To my parents

and my wife, Kelly,

for the unwavering support and patience that made this possible.

And finally, to Noah, my good luck charm and favorite source of motivation.

## Acknowledgements

For welcoming me into his lab and providing the necessary wisdom to guide me through this journey, I will be forever grateful to Dr. Jason Kahn, my thesis advisor and mentor. Thank you for your vision and effort in making this project and my professional development the success that it was.

# Table of Contents

Dedication.....	ii
Table of Contents.....	iv
List of Tables .....	viii
List of Figures.....	ix
List of Abbreviations .....	xii
Chapter 1: Introduction.....	1
1.1 DNA: The Genetic Polymer.....	2
1.2 DNA Topology: Maintaining Order Within a Cell.....	4
1.3 Balancing Supercoiling with Topoisomerase .....	8
1.4 Looping Proteins and Their Influence on Topology.....	10
1.5 Looping Proteins and Their Influence on Gene Regulation .....	11
1.6 Implications of Looping Size and Synthetic Manipulation .....	14
1.7 Incorporating Rigidity into a DNA Looping Protein.....	16
1.8 DNA Binding with Basic Leucine Zipper Proteins (bZip) .....	18
Chapter 2: The Design, Expression, and Purification of Artificial DNA Looping Proteins .....	25
2.1 The Coiled-Coil Rigid DNA Looping Protein.....	26
2.2 Design of a Tetrameric DNA Looping Protein.....	26
2.3 Design of the Dimeric Looping Protein.....	29
2.3.1 The <i>reverseGCN4</i> DNA Binding Protein.....	29
2.4 Expression of 4HB and LZD proteins .....	34

2.4.1	4HB Mutant Expression.....	34
2.4.2	LZD Mutant Expression .....	35
2.4.3	Extraction and Purification of 4HB Proteins .....	36
2.4.4	Concentration and Buffer Exchange into Storage Buffer .....	39
2.4.5	Circular Dichroism Analysis of LZD73 .....	41
Chapter 3: Binding Characterization of DNA Looping Proteins by EMSA.....		43
3.1	Overview.....	44
3.2	Materials and Methods.....	45
3.2.1	Binding/Ligation Buffer Formulation.....	45
3.2.2	Sample Preparation and Gel Analysis.....	46
3.3	Results.....	47
3.3.1	EMSA Analysis of 4HB mutants.....	47
3.3.2	EMSA Analysis of LZD Proteins .....	51
3.4	Discussion of Results.....	56
3.4.1	The 4HB Mutant Binding .....	56
3.4.2	The LZD Mutant Binding .....	57
Chapter 4: Length Dependent Loop Formation Using Ligase-Mediated DNA		
Dimerization .....		59
4.1	Overview.....	60
4.2	Materials and Methods.....	61
4.2.1	PCR Generation of Variable Length Fragments.....	61
4.2.2	Ligation Procedures .....	64
4.3	Results.....	65

4.4	Discussion of Results.....	68
Chapter 5: Topoisomer Product Distribution in Protein-Bound DNA Cyclization....		72
5.1	Overview.....	73
5.2	Principles Behind DNA Cyclization in Topology Studies .....	74
5.3	Materials and Methods.....	77
5.3.1	Design of Variable-Length DNA Constructs For Cyclization.....	77
5.3.2	Design of Vx 435-458 Binding Site Separation Fragments.....	80
5.3.3	Assembly, Radiolabeling, and Purification of DNA Constructs .....	82
5.3.4	Ligase-Mediated Cyclization of Protein-Mediated DNA Loops .....	83
5.3.5	EtOH Precipitation of Reacted Samples and Gel Analysis .....	84
5.4	Results.....	85
5.4.1	Formation of Topological Variants for Variable Length DNA Constructs Vx(153-448).....	85
5.4.2	Optimized Protein Concentration for Looping .....	88
5.4.3	Variable Length Vx(435-458) Formation of Topological Variants.....	90
5.5	Discussion of Results.....	95
5.5.1	Contribution of $\Delta$ Twist.....	95
5.5.2	Contribution of $\Delta$ Writhe .....	98
Chapter 6: Conclusions and Future Directions .....		100
6.1	Demonstration of an Artificial DNA Looping Protein .....	101
6.2	Further Characterization of LZD proteins .....	101
6.2.1	Identifying the Minimal Looping Length .....	101
6.2.2	Measurement of the Relative DNA Binding Angle Using FRET.....	102

6.3	General Future Directions.....	103
6.3.1	The Creation of Topological Domains using LZD proteins .....	103
6.3.2	Protein:DNA Nanostructures in Two and Three Dimensions .....	104
6.3.3	Introducing a Flexible Hinge into LZD .....	105
6.4	Conclusions.....	106
Appendix 1 - Sequences of Relevant DNA Constructs .....		108
Appendix 2: Additional Protein Purification Procedures .....		121
Bibliography .....		126

## List of Tables

Table 5.1 Vx(153-448) DNA fragments used in cyclization.....	78
Table 5.2 Vx(435-458) DNA fragments used in cyclization.....	81

## List of Figures

Figure 1.1 Figure Ideal B-DNA depicting the structure of the double helix .....	2
Figure 1.2 Plectonemic supercoiled DNA illustration.....	5
Figure 1.3 Supercoiled DNA depicting various degrees of supercoiling. ....	7
Figure 1.4 Repression levels of chromosomal <i>lacZ</i> expression.....	13
Figure 1.5 The structure of cortexillin. ....	18
Figure 1.6 A graphical representation of the residue interactions of GCN4. ....	20
Figure 1.7 The crystal structure of GCN4 bZip domain.....	22
Figure 1.8 The contact mapping between the $\alpha$ -helical region of the GCN4.....	23
Figure 2.1 Assembly of the tetrameric DNA looping design. ....	27
Figure 2.2 Schematic of tetrameric assembly of 4HB DNA looping proteins . ....	27
Figure 2.3 Modular assembly of the 4-helix bundle proteins .....	28
Figure 2.4 Modular assembly of <i>reverseGCN4</i> .....	30
Figure 2.5 GCN4 bZip and <i>reverseGCN4</i> fusion domains.....	31
Figure 2.6 Overlay of renderings for LZD73 and LZD87 proteins .....	32
Figure 2.7 The modular assembly of LZD73 and LZD 87 proteins. ....	33
Figure 2.8 Expression of the 4HB constructs. ....	36
Figure 2.9 Analysis of LZEE purification steps .....	38
Figure 2.10 Analysis of LZD 73 purification steps . ....	39
Figure 2.11 CD analysis of LZD73 protein. ....	41
Figure 3.1 EMSA of the 4HB construct with 23 bp DNA .....	48
Figure 3.2 EMSA of 145 bp DNA with GCN4 bZip control.....	49
Figure 3.3 EMSA of LZEE peptide and 177 mer DNA. ....	50



Figure 3.4 Asymmetrical sandwich complex assembly.....	51
Figure 3.5 EMSA of GCN4 bZip single binding control.....	53
Figure 3.6 EMSA of <i>reverse</i> GCN4 bZip single binding control .....	54
Figure 3.7 EMSA of LZD73 assymetric sandwich complex.....	55
Figure 4.1 Overview of dimerization bridged by sandwich complex formation .....	60
Figure 4.2 Plasmid diagram for Ix DNA fragments with Inv-2 binding sites. ....	62
Figure 4.3 Dimerization of Ix1 DNA.....	66
Figure 4.4 Dimerization of Ix2 DNA.....	67
Figure 4.5 Dimerization of Ix3 DNA .....	67
Figure 4.6 Sandwich complex-mediated dimerization kinetics.....	69
Figure 4.7 Proposed binding orientations for the sandwich complexes .....	70
Figure 5.1 Illustration depicting the conformational changes of cyclization.....	75
Figure 5.2 $\Delta$ Writhe and $\Delta$ Twist induced by protein looping.....	76
Figure 5.3 Plasmids used for producing the Vx(153-448) DNA fragments.....	79
Figure 5.4 Schematic of the ten plasmids used to construct the Vx(435-458) DNA..	80
Figure 5.5 Cartoon illustration depicting Vx(435-458) phased binding sites.....	81
Figure 5.6 Reaction outcomes that do not give novel topological products.....	86
Figure 5.7 Analysis of Vx153-448 cyclization constructs.....	87
Figure 5.8 LZD73 protein gradients .....	89
Figure 5.9 DNA-only cyclization of Vx(435-458). ....	91
Figure 5.10 DNA-only cyclization reactions for Vx(435-458) .....	92
Figure 5.11 Protein-induced distribution of topoisomers .....	93
Figure 5.12 Periodicity of topoisomer distributions for LZD73-mediated loops .....	94

Figure 5.13 Periodicity of topoisomer distributions for LZD87-mediated loops .....	94
Figure 5.14 Weighted average topoisomer distributions .....	99
Figure 6.1 FRET based measurement of binding site orientation .....	103
Figure 6.2 Topological domains formed by looping with LZD protein. ....	104
Figure 6.3 LZD Flex design.....	106

## List of Abbreviations

3C	chromosome conformation capture
4HB	four helix bundle
aa	amino acid(s)
AFM	atomic force microscopy
amp	ampicillin
ATP	adenosine tri-phosphate
B/L	binding/ligation (buffer)
bp	base pair(s)
bZip	basic leucine zipper
Cam	chloramphenicol
CD	circular dichroism
CIP	calf intestinal phosphatase
CV	column volume
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
EtOH	ethanol
IDT	Integrated DNA Technologies
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside

LacI	lactose repressor
Lk	linking number
LZD	leucine zipper dual-binding (protein)
MES	2-( <i>N</i> -morpholino)ethanesulfonic acid
MCS	multiple cloning site
NEB	New England Biolabs
PAGE	polyacrylamide gel electrophoresis
PNK	polynucleotide kinase
PCR	polymerase chain reaction
RNA	ribonucleic acid
sc	supercoiling
SDS	sodium dodecyl sulfate
ssDNA	single stranded DNA
TBE	Tris, borate, and EDTA (buffer)
Tw	twist
WLC	Worm-like Chain
Wr	writhe

## Chapter 1: Introduction

## 1.1 DNA: The Genetic Polymer

The elucidation of the double helix as the underlying structure behind nature's continuance symbolizes the birth of modern molecular biology and provided a new understanding of our genesis. By enabling an astonishing level of fidelity between generations, the semi-conservative method of replication would appear a logical extension of DNA's form. But this process requires the complete dissociation of the two helices, a task for which the structure of DNA is far from ideal. For starters, DNA is a very long molecule, narrow in width, and has a fairly short helical repeat (10.5bp), meaning it is heavily twisted. In a closed circle, such as a genome, pulling apart the strands for replication or transcription places immediate strain, in the form of over-twisting, on the remaining double stranded portion of the molecule.

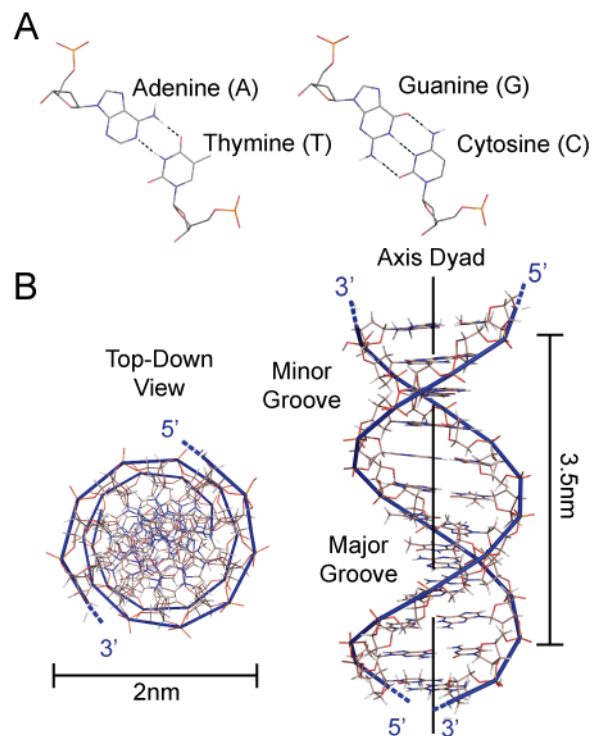


Figure 1.1 Figure Ideal B-DNA depicting the structure of the double helix. (A) Watson and Crick pair pairing and (B) two views of the double helix modeled from ideal B-DNA

The difficulties in separating the double helix over an entire genome were discussed by Watson and Crick almost immediately after their groundbreaking announcement of its structure (J. D. Watson & Crick, 1953a; 1953b). The double helix, a consequence of the conjunction of asymmetrical building blocks, demands a substantial amount of energy and protein regulation in maintaining the equipoise between being genetically accessible and structurally compact. Indeed, while proteins possess a remarkable tendency to mutate their shape, function, and relative size, DNA has remained nearly static in all physical aspects except for length. As organisms have grown in size and complexity over the eons, they have adapted to their burgeoning genome not by improving its underlying structure but rather by increasing and diversifying the proteins that organize and maintain it.

Indeed, while proteins possess a remarkable tendency to mutate their shape, function, and relative size, DNA has remained nearly static in all physical aspects except for length. As organisms have grown in size and complexity over the eons, they have adapted to their burgeoning genome not by improving its underlying structure but rather by increasing and diversifying the proteins that organize and maintain it. From histones or H-NS proteins that compact it to topoisomerases and gyrases that balance its strain, DNA is a highly regulated polymer that is ultimately under the control of proteins. Without a responsive and energetically demanding system to maintain this spatial organization, or topology of DNA, life could never have developed into the complexity observed today.

The advent of modern sequencing technology is delivering a wealth of data on the content of genomes across scores of species. The explosion of available

information has the potential to shower benefits on our civilization from the identification and elimination of genetic disorders to a unified theory of evolution. But the path from the genetic code to living organism is, like the molecule itself, hardly linear. The networks of genes and intricate feedback systems required for development demand coordination that is only beginning to be understood. There is a marked disconnect between the two-dimensional nature of genetic sequence and the three-dimensional life form to which it gives rise. Like all DNA, the human genome measures 2 nm in width but has a length that is orders of magnitude greater ( $10^8$  for *Homo sapiens*). That this molecule serves its function while compacted to fit inside a 6  $\mu\text{m}$  nucleus, attests to the complexity of its protein-regulated structure and underscores the need to comprehend the mechanisms behind its order. DNA structure, its topology, geometry, and geography, represent the foundation upon which genetic information is built, stored, and accessed. If we cannot observe, predict, and ultimately control the structure of DNA, the acquisition of its entire sequence will remain a feat of limited application.

## *1.2 DNA Topology: Maintaining Order Within a Cell*

The helical repeat of DNA, a direct property of the twisting nature of the double helix, dictates that, when in an aqueous environment, the two strands will cross one another roughly once every 10.5 bp. A second type of crossover event occurs when two separate double helix strands make a close approach at a node. This element of structure is referred to as writhe. As the molecule is compacted, the formation of these crossover nodes becomes increasingly common. Depending on the orientation of the crossover event, nodes may have either positive or negative quality.



As illustrated in Figure 1.2, the frequency and geometry of nodes result in the quantitative value of writhe. The amount of writhe reflects the degree of supercoiling, which is the underlying feature of DNA topology. This essential component of compaction was first described in the 1960's while studying the two structurally distinct forms of genetically identical polyoma virus DNA (Vinograd, Lebowitz, Radloff, Watson, & Laipis, 1965). But if these two identical sequences of DNA had different structural features, there must be a way to quantify the difference.

The means of quantifying the structural differences lies in the number of times the two strands cross each other through both helical repeat (the twist component) and

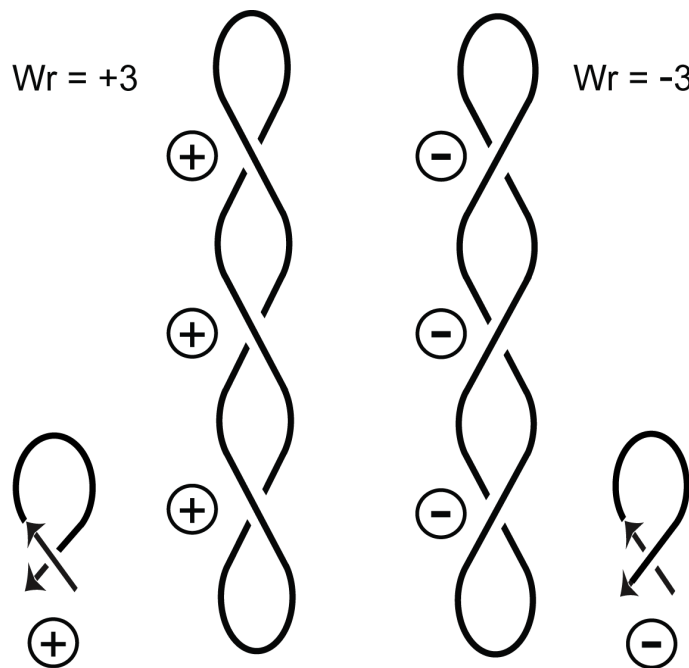


Figure 1.2 Plectonemic supercoiled DNA illustration. Each line represents double stranded DNA. The contribution of writhe in supercoiling is quantified by the formation of both (+) and (-) nodes leading to an increase or decrease in the linking number, respectively

through node formation (the writhe component). If two ends of a linear fragment of DNA are joined together in a closed circle, then the two strands of the double helix

are linked together by the number of times the strands cross, as per the helical repeat. This quantity must be an integer (as there are no partial crossovers in a closed circle) and represents the linking number of circular DNA lying in a plane. But fixing DNA to two dimensions is not an element of the real world. In fact, genomic DNA crosses over itself constantly in its natural environment. These crossover nodes are also linked in a closed circle of DNA and, as such, can be added to the number of helical repeat crossing events to provide an absolute linking number (Lk) for any given closed circle of DNA. DNA nodes, however, can have either positive or negative values depending on the orientation of the cross over.

As illustrated in Figure 1.2, a positive node increases the overall Lk value, while a negative change in writhe and an overall decrease in the linking number. Because the absolute value of Lk cannot change without breaking one or both strands of DNA, the linking number is an excellent means of quantifying DNA topology. As seen in Figure 1.3, plasmid DNA with populations that differ in their linking numbers can be easily resolved using agarose gel electrophoresis in the presence of an intercalating agent such as chloroquine. That the linking number remains unchanged ( $\Delta Lk = 0$ ) in a given closed circle of DNA, however, does not mean that the twist (Tw) and writhe (Wr) components remain static. The two elements can be readily inter-converted according to the following formula:

Eq.1  $\text{for } \Delta Lk = 0, \Delta Tw = -\Delta Wr$

This ability to relieve torsional stress by converting it to writhe is essential but clearly insufficient for dealing with the topological strain that arises during replication. To accommodate this systemic energetic barrier, the cell must employ a

means of changing the linking number such that overtwisting caused by the strand separation during replication can be relieved. If the strands could break then the change in either or both the twist and the writhe would result in a change in the linking number according to the following:

$$\text{Eq.2} \quad \Delta Lk = \Delta Tw + \Delta Wr$$

It was suggested in 1954, that cells may use an approach where one or both strands of the helix are broken so that torsional strain may be relieved through untwisting (Delbrück, 1954). Nearly two decades would have to pass before this theory could be validated when, in 1971, an enzyme termed the  $\omega$ -protein was isolated from *E. coli* (Wang, 1971). This enzyme, subsequently renamed DNA Topoisomerase I, possesses an ability to relax supercoiled DNA by nicking one strand and allowing it to rotate about the axis of the intact strand. Because this enzyme facilitated the breaking of one of the strands, the linking number could be changed.

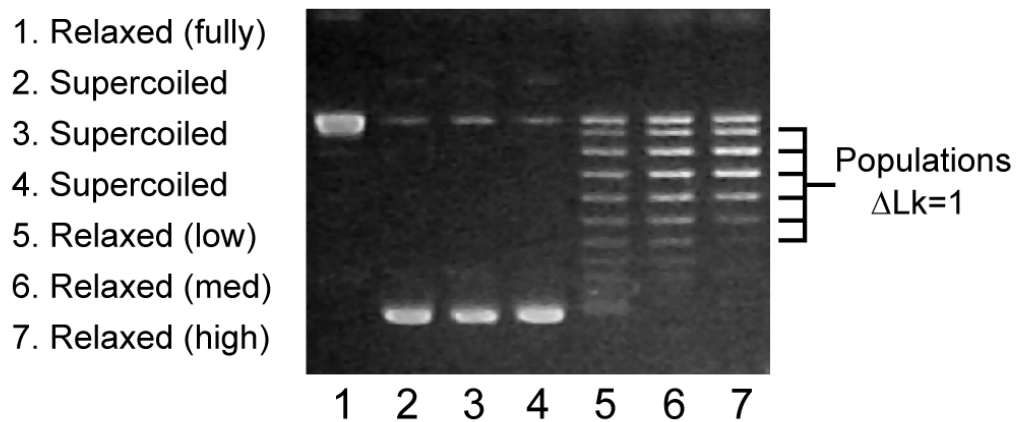


Figure 1.3 Supercoiled DNA depicting various degrees of supercoiling resolved on agarose gel with chloroquine. To form a distribution of topoisomer products, plasmid DNA was incubated with Topoisomerase I for an increasing amount of time (lanes 5,6,7). This gel is meant to illustrate how individual topoisomer populations can be resolved.

This was a monumental achievement for the nascent field of DNA topology and represented the first of a large and complex class of topoisomerase enzymes.

### *1.3 Balancing Supercoiling with Topoisomerase*

Though it is unsurprising that the topoisomerase class of enzymes exists, it is nonetheless fascinating to consider the many ways cells have evolved to maintain the topological balance throughout their genome. The immediate need for supercoiling is obvious, compaction, and nearly all cells maintain their genome as negatively supercoiled DNA (left-handed nodes). This topological state is maintained by the ATP-dependent enzyme DNA gyrase (Topoisomerase IIA) in bacterial and by histone wrapping in eukaryotes (Camerini-Otero & Felsenfeld, 1977; Gellert, Mizuuchi, O'Dea, & Nash, 1976). But chromosomal condensation is far from the only application of this structural phenomenon. For example, transcription factor binding has been shown, in some cases, to be dependent on the degree of negative supercoiling at the promoter site (Lamond, 1985). Furthermore, the opening of a transcription bubble by RNA polymerase II requires a degree of local untwisting and corresponding torsional strain that is compensated by the inherent negative writhe (Choder & Aloni, 1988). Though a preponderance of organisms maintain homeostasis with negatively supercoiled DNA, those living in extremely high temperatures, such as members of the *Sulfolobus* genus, have evolved a reverse gyrase, whose ATP-dependent activity introduces positive supercoiling (Kikuchi & Asai, 1984). While negatively supercoiled DNA aids in opening DNA for transcription, positively supercoiled DNA produces the opposite effect, thus

increasing the melting temperature to maintain genomic stability at very high temperatures.

The essential function and ubiquitous activity of topoisomerases has made them viable targets for cytotoxic drugs. Because DNA gyrase and the closely related Topoisomerase IV are both unique to the bacterial kingdom, inhibitors specific to their function, such as fluoroquinolones like Cipro, have been put to use as broad spectrum antibiotics (Maxwell & Lawson, 2003). Work on inhibiting eukaryotic topoisomerases has led to clinical applications in anti-cancer trials, as Topoisomerase activity is essential for replication (Hande, 1998). It is also possible that protein engineering work with Topoisomerases may prove useful in the future of genetic manipulation. One could see value in a Topoisomerase that possessed binding specificity that would limit its function to a predetermined location on the genome. In gene therapy, a targeted sequence may be histone-bound and inaccessible. A reverse gyrase enzyme that could target the region and induce positive supercoiling could aid in displacing the histones and allowing access to the area of interest. If we are to attain the ability to access and control genetic material on a level that stretches across the entire genome, topoisomerases may well play a pivotal role. However, for all their influence on DNA topology, the topoisomerase enzymes lack sequence specificity and thus act globally. In an event where topology must be controlled at a local level, such as the regulation of a specific gene, nature has adapted a second method of topological control, the DNA looping proteins. Protein-mediated loop formation provides a means of locking DNA in position. Manipulating DNA through

this approach offers specificity and reversibility and may serve an alternate platform to affect DNA structure by design.

#### *1.4 Looping Proteins and Their Influence on Topology*

The phosphate backbone of the double helix presents the molecule with several advantages within in a cell. The negative charge it carries contributes favorably to its solubility and makes its diffusion through cellular membrane unlikely. For proteins seeking to have some effect on DNA, this charge density serves as a beacon. It is not difficult to imagine how early peptides with dense regions of arginine and lysine could have first adapted to binding DNA. From transcription factors, to histones, to DNA repair enzymes, proteins have evolved to interact with DNA to perform a myriad of functions. As organisms evolved and their genomes expanded, proteins with DNA binding ability became increasingly valuable in the effort to maintain order.

Supercoiled DNA can be viewed as energetically primed. As discussed, it is easier to compact, transcribe, and replicate DNA that is negatively writhed. This energy is locked in position because the linking number of DNA cannot change unless one or both of the strands are broken. But DNA is not an infinitely stable molecule and the threat of single-strand nicking or double-strand breaks places the genome in structural peril. Fortunately, proteins have adapted to protect against these common threats by forming loops to lock DNA in position. DNA looping proteins are therefore able to create isolated regions of topology where the actions on regions are structurally separate from another. In *E. coli*, electron micrographs were able to observe such loops forming around a central core in the nucleoid (Kavenoff &

Bowen, 1976). This work, and others like it, led to the formation of the rosette theory to describe bacterial DNA structure. While still not fully understood, loop formation throughout the prokaryote genome is a highly regulated phenomenon, managed by a number of key proteins such as H-NS and HU (Noom, Navarre, Oshima, Wuite, & Dame, 2007; Thanbichler & Shapiro, 2006). Recently, it has been demonstrated that these topologically isolated domains can be achieved using natural looping proteins on engineered plasmids *in vitro* (Leng, Chen, & Dunlap, 2011). These examples of proteins exerting topological control on DNA suggest that manipulating DNA in an exact manner at specific sequences is quite possible. To date only natural looping proteins have been utilized to create topological domains using DNA engineered to incorporate specific binding sites. Expanding the engineering application to include modified or synthetic DNA looping proteins could vastly increase the scope of this application. With appropriate engineering, such proteins could be harnessed for work in gene therapy delivery systems or replication halting chemotherapy therapeutics.

### *1.5 Looping Proteins and Their Influence on Gene Regulation*

The compaction of DNA, a global event in principle, is managed, with few exceptions, by proteins that bind to DNA without regard for sequence recognition. Gene transcription, a process requiring access to a linear form of DNA, can be viewed as a local event and, in contrast, typically involves proteins that bind in a sequence specific manner. Because both of these extremes must coexist for survival, the genome is in constant state of balance between a need for compaction and a need for expansion. As discussed, the mechanisms employed to spatially manage DNA are impressive but relatively few in number. However, for the purpose of transcription,

the required specificity implicit in regulating thousands of unique genes has led to an immense diversity of control mechanisms. Leaving aside the discussion of signaling pathways that may add layers of complexity to gene regulation, the essence of transcription can be distilled to the notion of a genetic circuit, capable of being turned on or off.

Early insight into this regulatory approach came in 1961, from Jacob and Monod and their work with *E. coli*. They noticed that the expression of three proteins,  $\beta$ -Galactosidase, permease, and transacetylase was enhanced in the presence of lactose (Jacob & Monod, 1961). They theorized that the expression of the three genes, now known as *lacZ*, *lacY*, and *lacA* from the *lac* operon, were activated by lactose and repressed by some unknown agent in the absence of lactose. This agent was later identified as the lac repressor protein (LacI) whose own expression was coded by the *lacI* gene at the upstream portion of the *lac* operon. Its repression activity was linked to its ability to bind specifically to region of DNA within the *lac* operon, where it blocked RNA polymerase from binding (Gilbert & Maxam, 1973; Gilbert & Müller-Hill, 1966). Furthermore, the identification of two other local binding sites for LacI within the *lac* operon suggested possible DNA loop conformations *in vivo* and that these sites provided enhanced repression through cooperativity (Krämer et al., 1987; Oehler, Eismann, Krämer, & Müller-Hill, 1990). Looping was proven by a clever experiment that showed that repression levels of the regulated gene *lacZ* were dependent on the periodicity of the LacI binding sites (Bellomy, Mossing, & Record, 1988). This experiment was further refined and the limits of looping tested well below the 91 bp that separate the binding site in the wild



type operon (Müller, Oehler, & Müller-Hill, 1996). Figure 1.5, taken from Muller et al. demonstrates looping by correlating repression activity with the helical repeat of DNA. Amazingly, evidence of looping was observed at lengths down to 57.5 bp between operator sites.

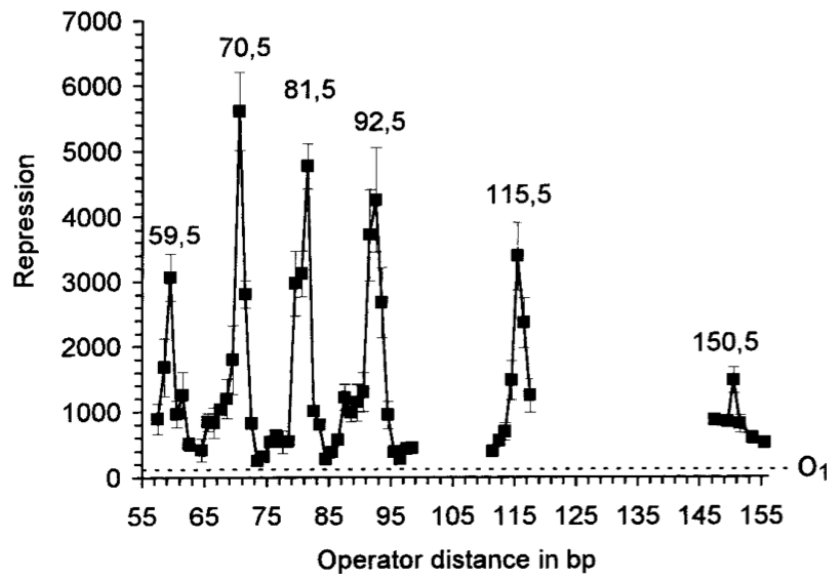


Figure 1.4 (From Müller et al., 1996) Repression levels of chromosomal *lacZ* expression with increasing spacing between the LacI operator sites. The repression is shown to be dependent on the phasing of the operators sites and correlates to the helical repeat of DNA presenting a classic demonstration of loop formation.

That looping existed and could occur at such small lengths led to an evolution of our understanding of the *lac* operon system. Its newly uncovered complexity confirmed DNA looping to be a means of enhancing the regulatory power of proteins involved in gene transcription.

While arguably the most characterized DNA looping protein, LacI is not alone in its mechanism. Another *E. coli* transcription pathway, the Gal repressosome utilizes looping and wrapping of DNA around the gal repressor protein (GalR) in its regulatory role (Haber & Adhya, 1988). This model is distinct from the *lac* operon in that a secondary protein, HU, is involved in binding and kinking DNA within the loop

thereby providing enhanced stability (Geanacopoulos, Vasmatzis, Zhurkin, & Adhya, 2001; Lewis, Geanacopoulos, & Adhya, 1999).

The relatively recent technique, chromosome conformation capture (3C), in which chromosomal DNA is covalently cross-linked to bound proteins and then those interactions are mapped by digestion, ligation, and PCR, has provided a systematic approach to DNA looping *in vivo* and has begun to elucidate its frequency (Davison et al., 2012; Tolhuis, Palstra, Splinter, Grosveld, & de Laat, 2002; K. Yun, So, Jash, & Im, 2009). The prevalence of looping in eukaryotes, and its capacity to exist over surprising large distances of tens or hundreds of kilobases, further underscores the significance of DNA looping as a means of spatial control within a cell.

### *1.6 Implications of Looping Size and Synthetic Manipulation*

DNA looping over very large lengths, such as those discovered using the 3C method, must overcome entropic hurdles to bring together these distant sites. The large lengths do mitigate the energetic cost of bending or twisting DNA, and it can be concluded that looping DNA many times longer than its persistence of 50 nm (roughly 150 bp) is independent of the geometry of the bound DNA (Hagerman, 1981). In contrast, looping events of much smaller scale, such as the 91 bp loop in the *lac* operon, require a far greater energetic cost as DNA become quite rigid at shorter lengths (Oehler et al., 1990; Shore & Baldwin, 1983a). The existence of looping well under the persistence length, such as the formerly mentioned LacI-mediated loop, has been explained, in part, by attributing a fraction of the energetic cost to flexibility inherent in the looping protein (Edelman, Cheong, & Kahn, 2003; Mehta & Kahn, 1999; Rutkauskas, Zhan, Matthews, Pavone, & Vanzi, 2009). If this

is the case, the ability of the protein to assume multiple conformations stabilized the small loop (Rutkauskas et al., 2009). The LacI protein, which is a tetrameric protein held together by a leucine-rich four-helix bundle (4HB), contains two regions of considerable flexibility: the hinge region separating the DNA binding domain from the N-terminal core domain and the proline-rich linker connecting the C-terminal core domain to the 4HB. Recent work involving DNA fragments with inherent topological strain induced by poly-adenine tracts (A-tracts), suggests that both an open and closed form of LacI may form depending on the contour of the DNA (Haeusler et al., 2012). In mutation studies involving the spacing of the LacI operator and its effect on repression rates, it was found that loops could form *in vivo* at lengths as short as 57 bp (Müller et al., 1996). Looping has been confirmed by the fact that repression levels depended on the periodic spacing of the operators and correlated to the helical repeat of DNA (Bellomy et al., 1988). This result is truly remarkable given that this represents distances slightly over one third the persistence length.

A competing theory of enhanced DNA flexibility at short lengths has been put forth to alternately explain the existence of very small loops. In this model, the formation of spontaneous kinks in DNA results in enhanced bending effects at short lengths. The theory was supported using DNA cyclization experiments of very short lengths (85-105 bp) where uni-molecular, or cyclized products formed with far higher frequency than predicted by common models used to describe DNA behavior such as the Worm-like Chain (WLC) model (Cloutier & Widom, 2004; 2005; Wiggins et al., 2006). The ratio of the formation of uni-molecular products and bimolecular product is expressed by the *j*-factor and has been used to determine the torsional rigidity of DNA

and calculate its persistence length (Shore & Baldwin, 1983b; 1983a). The spontaneous kink theory is currently a source of contention and the approach used to demonstrate it has been openly challenged (Du, Smith, Shiffeldrim, Vologodskaya, & Vologodskii, 2005). A DNA looping protein could be used to investigate this short sequence enhanced flexibility, but only if the protein served as a rigid link between the bound DNA. Naturally occurring looping proteins rely on inherent flexibility and/or additional DNA binding proteins to alter the loop topology and increase stability as seen in the *lac* operon and Gal repressosome (Becker, Kahn, & Maher, 2005; Roy et al., 2005). These natural adaptations result in such proteins being inapplicable for studying DNA flexibility in isolation. Lacking availability of a preexisting rigid DNA looping protein, our lab set out to engineer an artificial alternative.

### *1.7 Incorporating Rigidity into a DNA Looping Protein*

*De novo* protein design will, by definition, begin at the level of its building blocks. Because this protein must meet certain structural specifications, namely uniform rigidity, forethought must go into how the amino acid sequence will ultimately fold. Of the limited secondary structures observed in peptide folding, it seemed logical to commence with a comparison of their relative flexibility. While no organic polymer with cellular origins can be considered truly rigid, as compared to macroscopic things such as lumber and steel, the relative stiffness of microscopic polymers can be rated using metrics such as persistence length. The persistence length can be thought of as a way of expressing the energy required to bend a

polymer. As seen in equation 3, the free energy of bending is directly correlated to the persistence length,  $a$ , over the contour,  $L$ , with a total bend angle,  $\Delta\Theta$  (Kahn & Crothers, 1998):

Eq. 3 
$$\Delta G = \frac{aRT}{2L}(\Delta\Theta)^2$$

Molecular-dynamics simulations performed on peptides that consisted of a continuous  $\alpha$ -helix concluded the structure to have a persistence length of 100 nm, or twice that of DNA (Choe & Sun, 2005). Furthermore, similar analysis on the structure of a coiled-coil of  $\alpha$ -helices, like that in the leucine zipper motif, increased the persistence length to nearly 150 nm (Wolgemuth & Sun, 2006). In contrast, the alternative secondary structure,  $\beta$ -sheets, in both parallel and anti-parallel form, were computationally shown to be significantly more flexible, with the frequent turns facilitating bending deformations (Choe & Sun, 2007; Emberly, Mukhopadhyay, Tang, & Wingreen, 2004). Random coil secondary structure was not considered for our application. The leucine zipper motif is a well-characterized coiled-coil structure of  $\alpha$ -helices. Of the natural structures available to serve as a template for our initial design, it was believed to offer the greatest potential for incorporating rigidity into a DNA binding protein.

Cortexillin is an actin-bundling protein in *Dictyostelium discoideum* that plays a major role in cellular shape, chemotaxis, and chromosome separation (Faix et al., 1996; Gerisch, Faix, Köhler, & Müller-Taubenberger, 2004). One of its most interesting features, however, is structural; it dimerizes through the formation of a

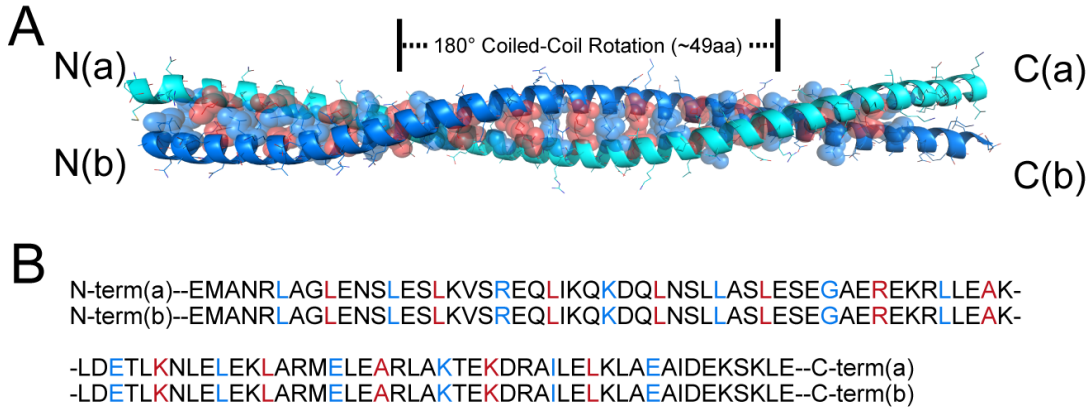


Figure 1.5 (A) The 101 aa structure of cortexillin in dimer form is a continuous coiled-coil motif. (B) The sequence highlights the hydrophobic elements of the a' and d' positions of the helical repeat, blue and red, respectively (also shown in space filling form in (A)). Image produced using Pymol, structure reference PDB:1D7M.

This region has been frequently used to study coiled-coil structure and played a major role in deciphering the amino acid trigger-sequence that dictates the oligomerization state in coiled-coil structures of two or more helices (Ciani et al., 2010). Figure 1.5 was generated using the crystal structure solved by Burkard and colleagues and illustrates the large coiled-coil feature of cortexillin (Burkhard, Kammerer, Steinmetz, Bourenkov, & Aebersold, 2000). Like nearly all coiled-coil dimers, cortexillin associates in a parallel orientation and displays a left-handed geometry along the helical axis. The crystal structure has been used to calculate a rotational period of roughly 49 aa (or 7 heptad repeats) for every 180° of twist. This rotational feature was taken into consideration when designing the length of our looping proteins and its effect on binding site orientation.

### 1.8 DNA Binding with Basic Leucine Zipper Proteins (bZip)

The bZip structural motif is a DNA binding domain used in a class of transcription factors whose origins have been traced back one billion years

(Amoutzias et al., 2007). Because the leucine zipper is a coiled-coil structure, use of a bZip DNA binding domain is appealing in the design of a rigid DNA looping protein. In an effort to minimize the potential for flexibility, the peptide structure should be continuous in nature, meaning that the coiled-coil motif is to be maintained for all, or nearly all of the structure. c-Myc is a DNA binding protein found in humans that was first identified by way of its sequence similarity with the oncogene v-Myc from the avian myelocytomatosis virus (Dalla-Favera et al., 1982). Structurally this protein is significant because its similarity to CCAAT-enhancer binding protein (C/EBP), specifically the placement of leucine residues at the *d* position of the heptad repeat (*abcdefg*) over the span of four helical repeats, led to the discovery of the leucine zipper motif and its recurrent association with DNA binding regions (Landschulz, Johnson, & McKnight, 1988). Further characterization of the structure uncovered the importance of the electrostatic interactions between the *e* and *g'* residues between helices in providing stability and dimerization specificity (O'Shea, Lumb, & Kim, 1993; O'Shea, Rutkowski, & Kim, 1992). As seen in Figure 1.6, the dimerization of the GCN4 homodimer is stabilized by the hydrophobic interactions of the *a* and *d* residues of one  $\alpha$ -helix with the *a'* and *d'* residues of its pairing  $\alpha$ -helix. Additionally, electrostatic interactions of the *e* residues of one helix with the *g'* residues of the helix lead to greater stability. To the N-terminal of the leucine zipper, the DNA binding region of this motif makes frequent use of the basic amino acids lysine and arginine as contact points with the DNA phosphate backbone. It is the combination of a basic binding site and the leucine zipper that has led this to this

broad class of DNA binding proteins being referred to as the basic leucine zipper, or the bZip family.

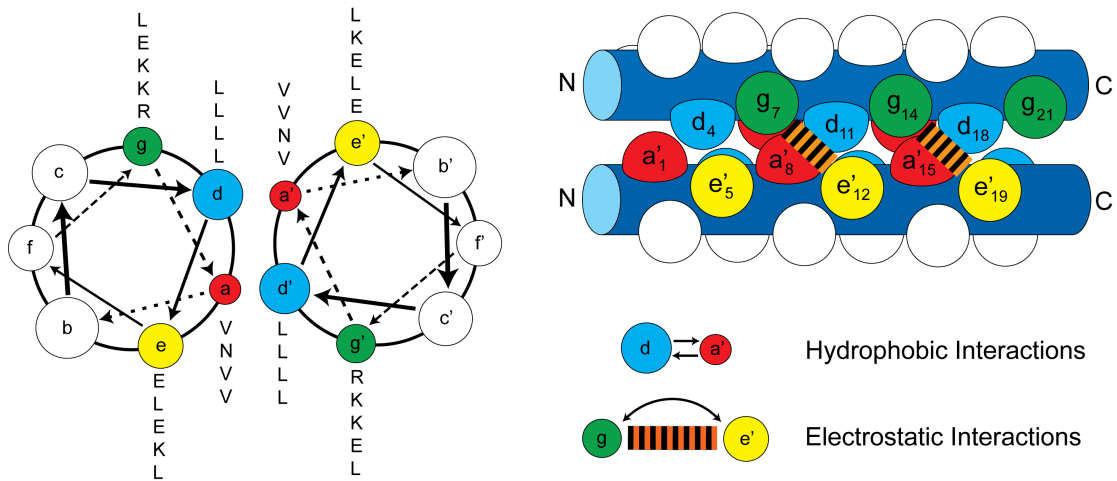


Figure 1.6 A graphical representation of the residue interactions of the GCN4 leucine zipper. Left, an  $\alpha$ -helix diagram depicting the hydrophobic burying of the a and d residues in the coiled-coil. Right, a space filling illustration showing both the hydrophobic burying of the a and d (red and blue spheres) as well as the interaction between the g and e' residues between alpha helices (green and yellow spheres).

There exists a great deal of variety among the bZip members. All are capable of dimerization but many, such as the human fos/jun pair as heterodimers (Abate, Luk, Gentz, Rauscher, & Curran, 1990), while others such as the yeast factor GCN4 form homodimers (Ellenberger, Brandl, Struhl, & Harrison, 1992; O'Shea, Rutkowski, Stafford, & Kim, 1989). Among the DNA binding regions there also exists a degree of structural variance. Previous work with c-Myc suggested that it was capable of forming a tetramer that could bind DNA at two points to form a loop (Ferré-D'Amaré, Pognonec, Roeder, & Burley, 1994). While the rigidity of such a structure was unknown, it presented an interesting approach to how coiled-coils could incorporate two DNA binding regions. c-Myc, as a rigid looping protein, had several shortcomings. Biochemical analyses indicated c-Myc bound folded into a Helix-



Loop-Helix motif for the transition from the helical binding region to the helical zipper domain (Fisher, Parent, & Sharp, 1993). This structural feature was then demonstrated in the solved crystal structure as a heterodimer with its protein counterpart Max (Nair & Burley, 2003). Here, the loop region junction likely plays a role in stabilizing the interaction and enhances the binding but may afford the protein flexibility and as such should be avoided in our design. Moreover, recent work with c-Myc and its sometimes dimerization partner Max demonstrated that while the proteins could fold in a manner that allowed for binding two strands of DNA, in a structure termed a “sandwich complex”, the binding was found to be too weak to support the formation of a DNA loop (Lebel, McDuff, Lavigne, & Grandbois, 2007). This prior work would exclude c-Myc from further consideration in the design process, but it was illuminating in suggesting a route to combine two DNA binding sites along a coiled-coil motif.

The yeast transcription factor GCN4 was identified by its association with the His3 gene and its role in regulating amino acid biosynthesis during periods of starvation (Hope & Struhl, 1985). Further analysis indicated that it bound to DNA in dimeric form (Hope & Struhl, 1987). The following year, the c-Myc & C/EBP correlation led to the announcement of the bZip family motif and it was quickly noted that the DNA binding region of GCN4 aligned with this proposed structure. The structure of the leucine zipper region of the protein was then solved in 1991, which solidified its status in the bZip family (O'Shea, Klemm, Kim, & Alber, 1991). The complete bZip domain bound to the pseudo-palindromic AP1 DNA (5'-ATGACTCAT-3') was solved the following year by Ellenberger, et al. and revealed a

continuous stretch of  $\alpha$ -helices extending from the coiled-coil region straight through the DNA-binding site (Ellenberger et al., 1992). An additional structure (depicted in Figure 1.7), solved by Tom Richmond's group, shows GCN4 bound to the palindromic CREB DNA (5'-ATGACGTCAT-3'). This structure was solved to a higher resolution enabling accurate characterization of the Protein:DNA contact

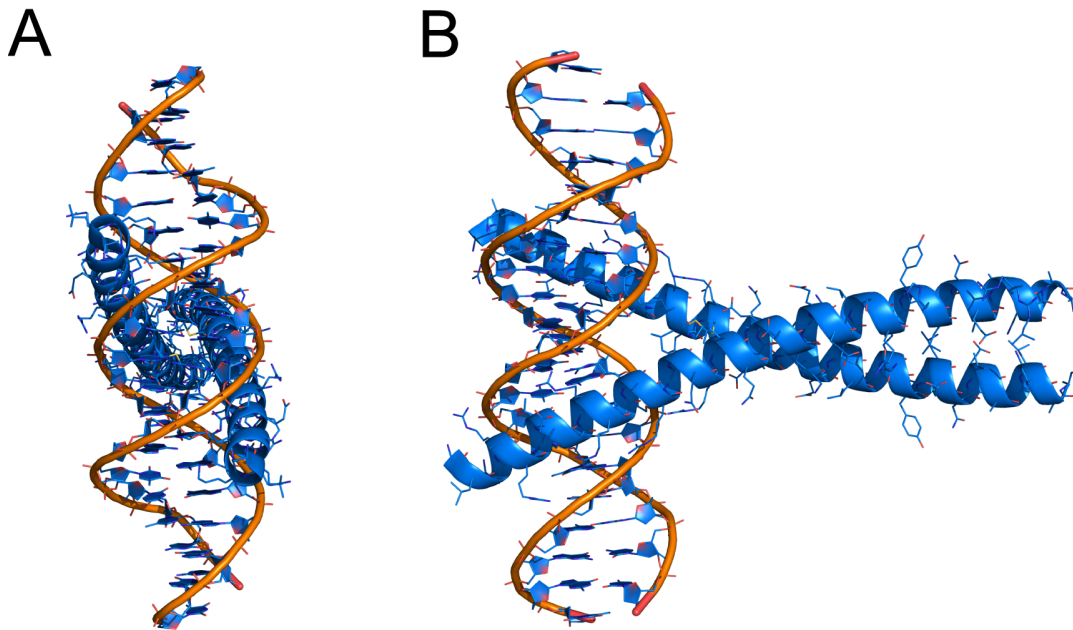


Figure 1.7 The crystal structure of GCN4 bZip domain illustrates a continuous  $\alpha$ -helical structure between the coiled-coil and the DNA binding site. The continuous  $\alpha$ -helix is intended to confer rigidity to the proteins. Image created using Pymol with PDB:1DGC

points (Keller, König, & Richmond, 1995). This work was able to provide a contact map fully elucidating the interaction between one of the  $\alpha$ -helices and half of the palindromic binding sites. This is depicted in Figure 1.8, taken from Keller, et al. 1995. The continuous extension of  $\alpha$ -helical structure between the coiled-coil region and the basic DNA binding site is of particular interest because this structure confers the greatest chance of maintaining rigidity if applied to a DNA looping protein.

GCN4 was, therefore, selected as the starting template for our artificial DNA looping protein. For a means of combining two DNA binding-sites our design turned elsewhere.

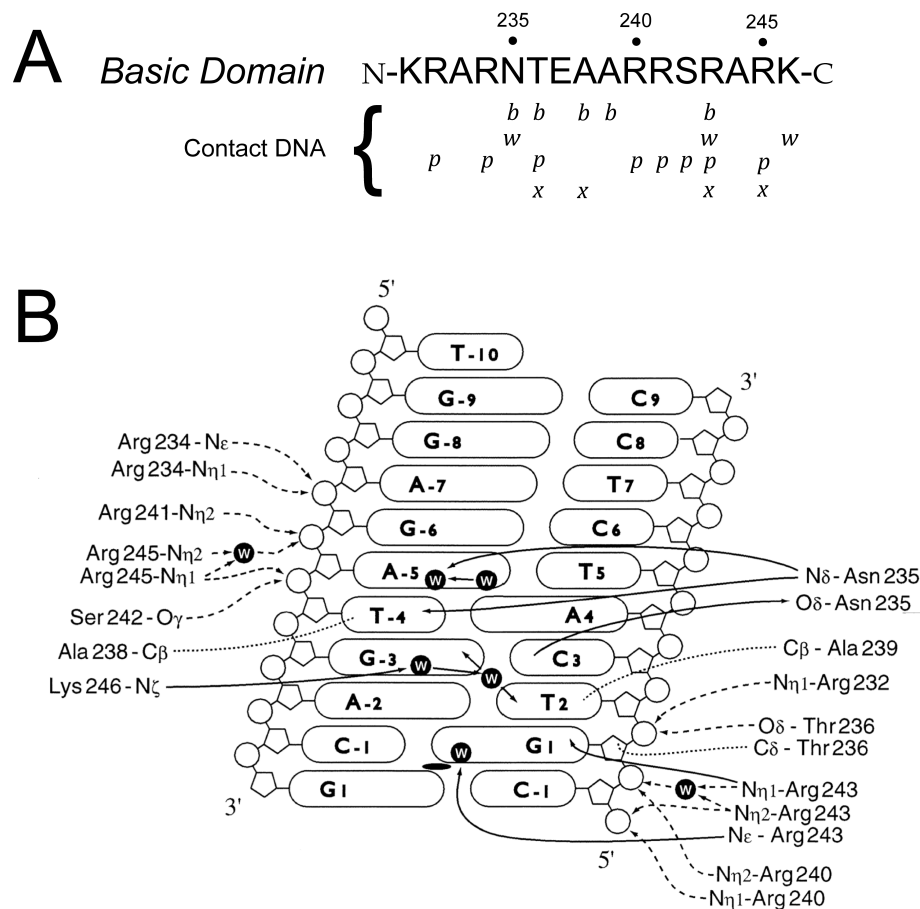


Figure 1.8 (From Keller et al, 1995) The contact mapping between the  $\alpha$ -helical region of the GCN4 monomer and the CREB site DNA (half-binding site). (A) a grid depiction of amino acids forming bonds with DNA base pairs (b-direct to base, w-through water to base, p-direct to phosphate backbone, x-via water to phosphate backbone). (B) Visual contact map of these bonds.

Increasingly, engineers have looked to biomimetics to provide solutions to medical challenges from tissue regeneration to gene delivery (Chae et al., 2011; Coburn et al., 2011). For applications that require *in vivo* DNA manipulation, protein-mediated DNA looping is an apt target for such modeling. By offering a broad range

of size possibilities and the incorporation of highly specific sequence recognition, such a system offers tremendous potential for eliciting control over DNA. It will undoubtedly take a great deal of bioengineering to convert a looping concept into a clinical reality, but it can begin with a simple statement of purpose: design an artificial DNA looping protein and investigate how it can manipulate DNA structure. This thesis describes the design, purification, and expression of a series of artificial proteins (Chapter 2) the binding characterization of the various peptides (Chapter 3), evidence of transient DNA loop formation (Chapter 4), and subsequent analysis of the topological manipulation induced by loop formation with our proteins (Chapter 5). By creating an artificial DNA looping protein, we have created a platform for affecting DNA topology by design. Additionally, the binding-site specificity and ability of the protein to alter the DNA binding site orientation through design modifications makes this work potentially well suited to developing self-assembling protein:DNA nanostructures.

## Chapter 2: The Design, Expression, and Purification of Artificial DNA Looping Proteins

## *2.1 The Coiled-Coil Rigid DNA Looping Protein*

The argument for using the coiled-coil structure in designing a rigid DNA looping protein is presented in sections 1.7 and 1.8. The application of this concept resulted in two major design approaches: a tetrameric design and dimeric design. Both of these structures would be assembled using homodimers with GCN4 DNA binding domains. Future work with this project may find the use of hetero-multimeric assembly appealing, as this would provide greater variety to the DNA binding sequence, which in our design is limited to palindromic sequences. Such a design was not considered in our application here. This chapter will describe the design and synthesis of the tetrameric and dimeric DNA looping protein designs used in this project.

## *2.2 Design of a Tetrameric DNA Looping Protein*

The LacI DNA looping protein folds into a stable tetramer as a dimer of dimers, in which dimeric core domains are held together by a leucine-rich four-helix bundle (Alberti, Oehler, Wilcken-Bergmann, & Müller-Hill, 1993; Alberti, Oehler, Wilcken-Bergmann, Krämer, & Müller-Hill, 1991). Crystal structure analysis of the core and tetramerization domains revealed that the 4HB domain, in which the helices are arranged in an anti-parallel orientation, positions the N-terminal ends to lie slightly farther apart than the helices found in leucine zipper dimers (Friedman, Fischmann, & Steitz, 1995). Structural studies involving the folding of leucine heptad structures into dimer, trimer, or tetramer products revealed that very subtle sequence changes

could interconvert the ultimate oligomeric state (Betz, Liebman, & DeGrado, 1997; Noom, Navarre, Oshima, Wuite, & Dame, 2007; Oakley & Hollenbeck, 2001; Thanbichler & Shapiro, 2006). These observations combined with the earlier suggestion that c-Myc/Max could form a tetramer, albeit an unstable one, lead us to design a bridge to combine two GCN4 bZip regions with a LacI 4HB.

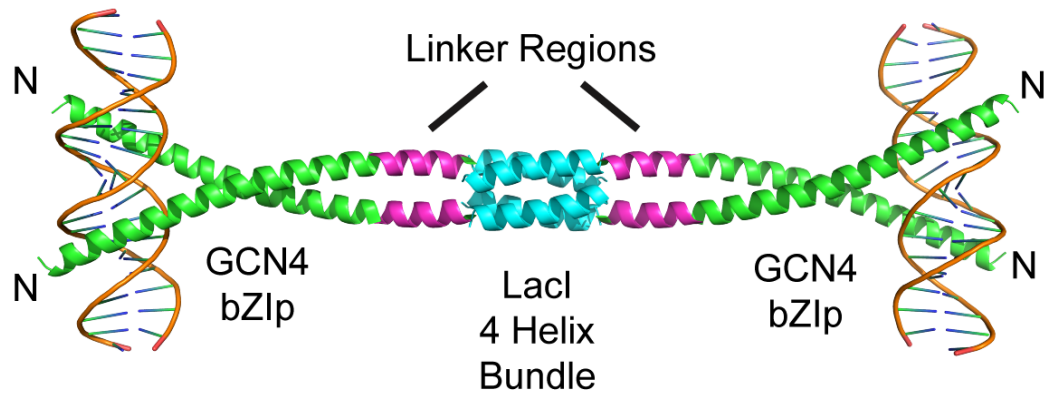


Figure 2.1 Assembly of the tetrameric DNA looping design with two GCN4 DNA binding domains (green) fused with the LacI tetramerization domain (cyan) and incorporating a short linker sequence (magenta) to preserve the  $\alpha$ -helical repeat.

Figure 2.1 illustrates the putative assembly of the designed tetrameric looping protein.

The inability to exactly fit the junction between the 4HB domain (cyan) and the

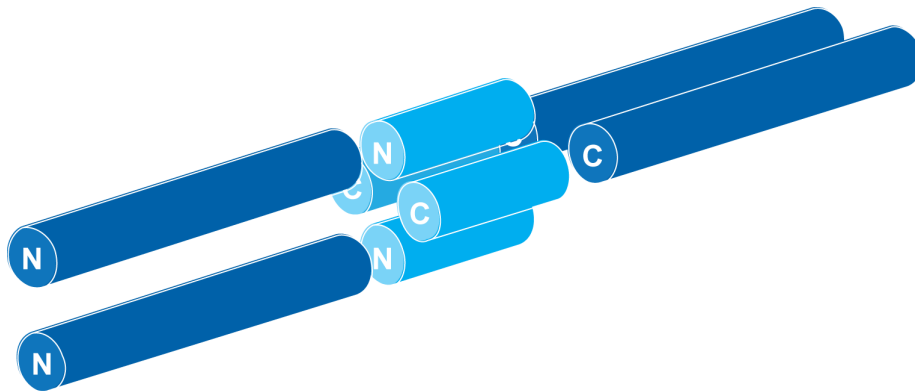


Figure 2.2 Schematic illustrating the assembly of the tetrameric DNA looping proteins using the LacI 4-helix bundle (light blue) and the GCN4 dimeric bZip domain (dark blue). The kitty-corner positioning of the 4-helix bundle provides a transition from tetrameric to dimeric structure.

GCN4 coiled-coil (green) was addressed by incorporating a heptad repeat linker - (magenta) to allow the coiled-coil helices to partially separate as they transitioned to the 4HB. Figure 2.2 is a schematic representing the assembly of the tetrameric DNA looping proteins.

The dense packing of hydrophobic residues in an extended leucine zipper may present solubility issue for our peptides. To account for the possibility of an insoluble product and the unknown element of transitioning between a coiled-coil and 4HB domain, four mutants were designed where each incorporated a unique linker. Genes expressing these four mutants were synthesized and cloned into plasmid pRSETA by

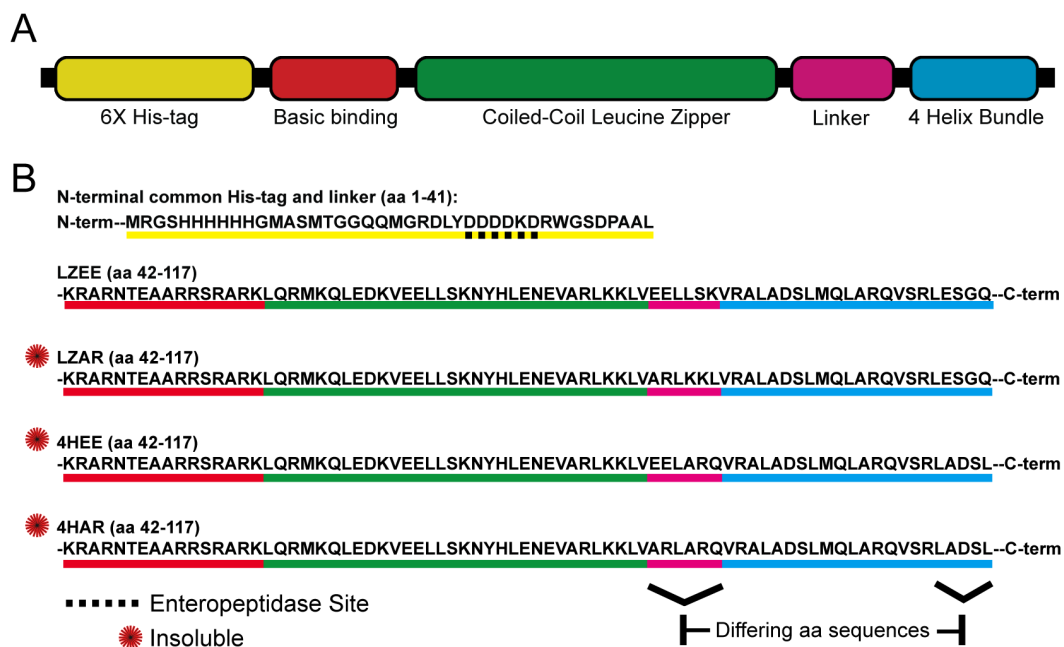


Figure 2.3 Modular assembly of the 4-helix bundle (4HB) proteins (A). Sequences given for the 4 constructs with the various domains underlined according to purpose: yellow – common N-terminal 6X histag and Enteropeptidase site (dashed underline), red – basic binding region, green – leucine zipper region, magenta – linker, blue – 4 helix bundle region.

Jason Kahn, expressed and purified as described in section 2.4. Figure 2.3 illustrates the modular design of the four constructs LZEE, LZAR, 4HEE, and 4HAR.



### 2.3 *Design of the Dimeric Looping Protein*

As indicated in Figure 2.3, three of the four tetrameric constructs expressed as insoluble peptides. This conclusion is taken from SDS PAGE analysis of the soluble lysis and insoluble pellet done during purification (Figure 2.8). While purification of these peptides was achievable using 6 M guanidine, efforts to refold the proteins upon removal of the guanidine proved unsuccessful. Additionally, binding analysis of the soluble LZEE construct provided evidence that the protein was not folding into a tetrameric state capable of binding two DNA fragments (see section 3.2.1).

It was thereby necessary to develop a second approach to designing an artificial looping protein. This subsequent engineering effort was more an extension of the previous design rather than a complete restructuring. The arguments for the coiled-coil motif conferring rigidity were sound and the strong binding of the GCN4 basic binding site had no shortcomings. The problem resided with the tetrameric domain and the likely possibility that dimerization rather than tetramerization of LZEE resulted in a more stable structure. Instead of a tetrameric linking domain we turned to a simpler assembly, a dimeric leucine zipper dual-binding (LZD) protein.

#### 2.3.1 The *reverse*GCN4 DNA Binding Protein

The inspiration for the next step came from work on the GCN4 peptide by Martha Oakley. Her group's investigation into the folding of bZip peptides led her to ask whether there was an inherent thermodynamic reason that all bZip DNA binding proteins position the basic region to the N-terminal side of the leucine zipper domain (Hollenbeck, Gurnon, Fazio, Carlson, & Oakley, 2001). In an experiment that can

only be described as essential to this project, her lab reconstructed the GCN4 peptide by inverting the order of the two domains and positioning the binding region at the C-terminal of the peptide, as illustrated in Figure 2.4.

### GCN4 bZip



### *reverse*GCN4 bZip



Figure 2.4 Modular assembly of *reverse*GCN4 created by Hollenbeck and Oakley (2001). The reversal of positions of the basic binding region (yellow) and the leucine zipper region (green) was performed to access whether there was a thermodynamic reason for the evolution of the N-terminal basic region arrangement among natural bZip DNA binding proteins.

The protein was simply named *reverse*GCN4 or rGCN4. To avoid confusion with recombinant nomenclature, it will only be referred to here as *reverse*GCN4.

Empirical work with the  $\alpha$ -helical phasing of the basic regions with respect to the leucine zipper using binding assays involving DNA with variants of an inverted CREB site produced a peptide that could bind DNA with near wild-type affinity ( $K_d = 29$  nM). The mutated binding site sequence Inv-2 (5'-GTCATATGAC-3') resulted in the highest affinity and was a perfect inversion of the GCN4 specific CREB site (5'-ATGACGTCAT-3'). The successful protein mutant utilized a 7 aa linker (-LQKLQRV-) between the GCN4 leucine zipper and the now C-terminal basic binding domain. The fusion of these two domains without disrupting the DNA

binding regions, preserves the previously determined DNA interaction maps. A model of the *reverse*GCN4 peptide is depicted in Figure 2.5 along with the GCN4 bZip peptide.

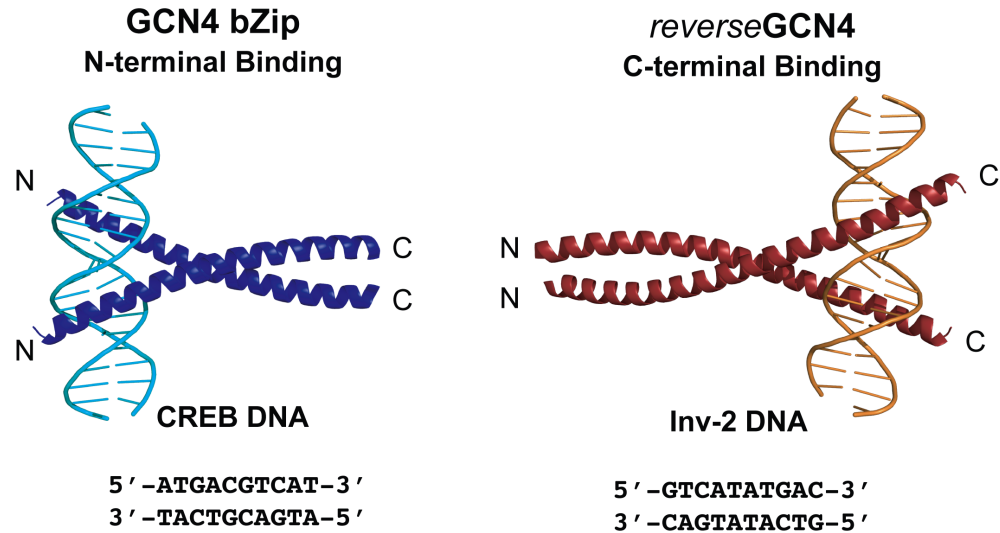


Figure 2.5 The two DNA binding domains that will be fused to form the leucine zipper dual-binding (LZD) protein. Left, the N-terminal domain binds specifically to CREB DNA (from Keller et al. 1995 – rendered in Pymol PDB:1DGC). Right, the C-terminal domain binds specifically to Inv-2 (Inverted CREB) DNA (image is a Pymol generated illustration).

This design should not be confused with work that reverses the sequence of amino acids from C to N-terminal. This structural change has previously been done with the leucine zipper sequence of GCN4 in creating a *retro*GCN4 peptide, which folds into a stable 4-helix bundle (Mittl et al., 2000).

The *reverse*GCN4 artificial protein presented a perfect opportunity to simplify our looping protein into a dimeric structure. By fusing the GCN4 bZip peptide with the *reverse*GCN4 peptide sequence the folded dimer should contain two DNA binding domains. The amino acid sequence separating the two binding sites was determined by aligning the *reverse*GCN4 sequence with GCN4 bZip resulting in a 73 amino acid sequence from this beginning of the N-terminal binding site to the end of

the C-terminal binding site. The protein design was termed LZD73. A gene expressing this peptide was cloned into pRSETA that incorporated an N-terminal 6X his-tag and Enteropeptidase cleavage site (-DDDKD-). The left-handed geometry of the coiled-coil motif presented a unique opportunity to adjust the angles between the two DNA strands. Because the coiled-coil wraps around itself and the binding site of the DNA is perpendicular to the coiled-coil axis, an extension of the coiled-coil should result in a change in the relative binding. To investigate this possibility, a second looping protein mutant was designed to incorporate an additional 14 amino acids between the GCN4 leucine zipper and the *reverse*GCN4 linking sequence. Keeping with the nomenclature established with LZD73, the additional 14 amino acids is reflected in the name LZD87. An N-terminal overlap of models for LZD 73 and LZD87 bound to CREB and Inv-2 DNA is depicted Figure 2.6.

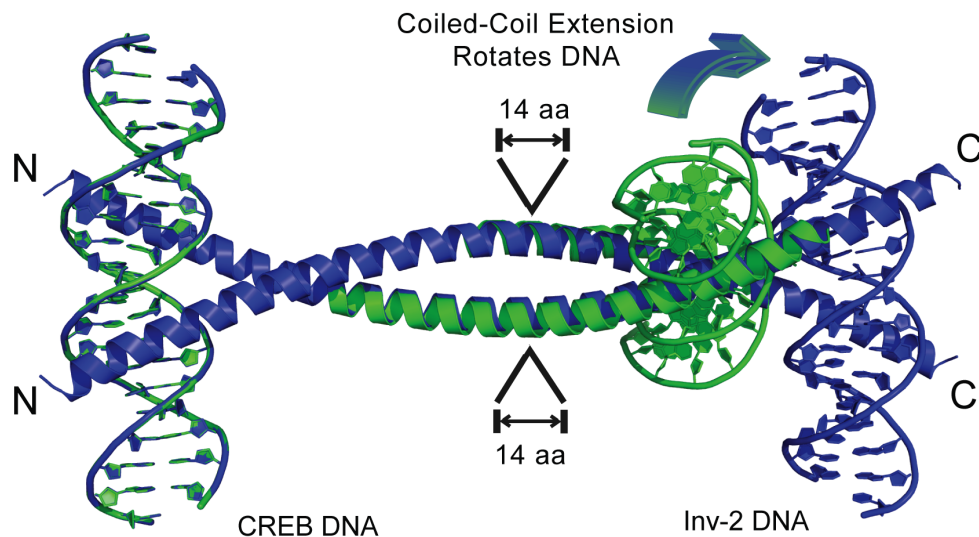


Figure 2.6 Overlay of renderings for LZD73 (green) and LZD87 (blue) DNA binding proteins bound to 20 bp DNA with either CREB or Inv-2 site sequence at the N-terminal and C-terminal, respectively. Pymol image illustrates the coiled-coil left-handed orientation and how the length change has leads to a rotation of the relative binding sites

The effects of the addition of 14 amino acids can be seen in the change in binding orientation of bound DNA segments.

Figure 2.7A illustrates the modular assembly of these two genes and 2.7B lists the amino acid sequence for each. By extending the leucine zipper domain by two heptad repeats, the hydrophobic content of the peptide was increased.

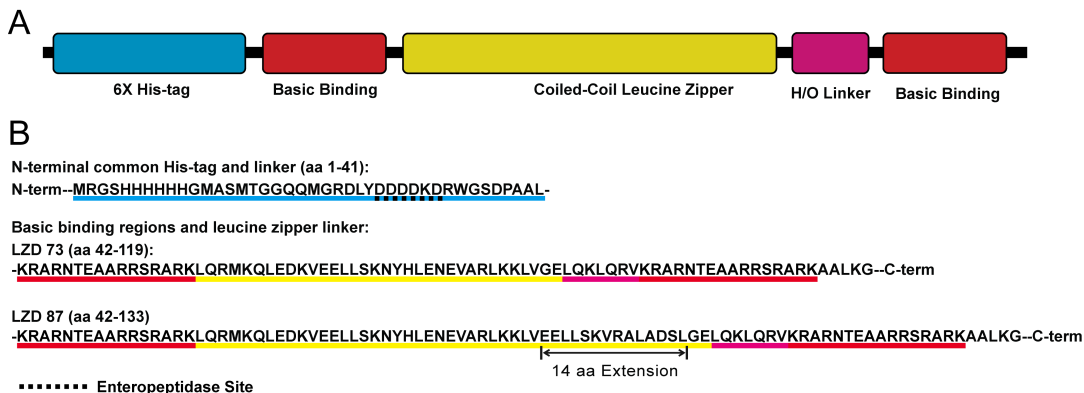


Figure 2.7 The modular assembly of LZD73 and LZD 87 proteins. (A) A depiction of the fusion of two basic DNA binding domains by a continuous coiled-coil domain and the necessary C-terminal linker region (H/O linker) determined in Hollenbeck and Oakley, 2001. (B) Sequences used in the design with the underlined regions corresponding to the modular illustration depicted in (A).

The solubility problems encountered in the 4HB mutant work raised concerns that this might lead to similar folding difficulties. In order to maximize the likelihood that this mutant would be soluble, the 14 aa sequence was taken directly in frame from LZEE, the soluble 4HB peptide. For visualization purposes, two models were generated using Pymol (see Figure 2.6). This image is meant to be illustrative and does not reflect any knowledge of the actual binding site angle orientation. In the figure above, the N-terminals have been aligned to highlight the binding site orientation differences at the C-terminal domain.

## 2.4 *Expression of 4HB and LZD proteins*

All reagents were purchased from Fisher Scientific with the exception of [ $\gamma$ - $^{32}$ P]-ATP, which was purchased from Perkin Elmer. Polynucleotide kinase was purchased from New England Biolabs (NEB). Protein chromatography was performed on the AKTA FPLC using columns purchased from GE Healthcare. Centrifugal filters were purchased from Millipore. Bio-spin 6 columns were purchased from Bio-Rad.

### 2.4.1 4HB Mutant Expression

Each of the four 4HB sequences denoted previously were prepared by oligonucleotide synthesis and mutually primed extension to give the plasmids pLZEE, pLZAR, p4HEE, and p4HAR. The expressed sequence contained an N-terminal 6X histidine tag for metal chelate affinity purification as well as an Enteropeptidase binding/cleavage sequence (-DDDDKD-) between the his tag and the 4HB open reading frame. The plasmids were transformed into electrocompetent BL21 DE3 (pLysS) cells by electroporation. The ORF sequence for each of these proteins is found in Appendix A. After rescue with SOC (1 mL) and 1 hr at 37 °C with shaking, the cells (15  $\mu$ L) were streaked on LB agar containing ampicillin (100 mg/L) and chloramphenicol (40 mg/L). The plates were then incubated overnight at 37 °C. A single colony was selected the following day and expanded overnight in a 5 mL LB culture (+Amp/+Cam) with agitation, at 37 °C. The culture was then used to inoculate a pre-warmed 1 L LB (+Amp/+Cam again) solution in a 4 L Erlenmeyer flask in the morning and allowed to grow for 4-6 hours until the optical density (OD<sub>600</sub>) reached 0.6. Expression was induced by the addition of IPTG (0.5 mM) and

the cells were allowed to express for 3 hours. The cells were harvested by centrifugation for 15 minutes at 12,000 x g and the wet pellets were frozen and stored at -80 °C unless purification was immediately implemented. Typical yields for 1 L harvests in this procedure were 2.0-2.5 g cell paste (wet).

#### 2.4.2 LZD Mutant Expression

Plasmids containing the sequences coding for LZD73, LZD87, and the single binding C-terminal control *reverseGCN4* were transformed into *E. coli* BL21 DE3 (pLysS) cells, selected for expansion and then grown in 5 mL starter culture as described for the 4HB mutants. Because of slower growth relative to the previous mutants, the timescale for pre-induction growth and expression length was adjusted accordingly to maximize yield. This retarded growth for cells carrying the LZD protein genes is likely due to leaky expression of the high-copy pRSETA expression system. It can be inferred that the LZD proteins are toxic for the host cells. It is possible that use of pLysE in place of pLysS could increase the growth rate during the pre-induction stage. Relative to pLysS, pLysE has a higher expression of T7 lysozyme, which binds to and inhibits T7 RNA polymerase. The basal expression of T7 RNA polymerase during pre-induction growth leads to leaky expression of the target pRSETA-based gene, and the leaky expression of a toxic protein is the likely cause of the diminished growth rate. For pre-induction growth, the 5 mL starter culture was used to inoculate 1 L of 37 °C LB that had been pre-warmed overnight. This step is performed early in the morning, because growth is very slow at this step. After 10 hours of shaking at 37 °C, the cells typically have reached an OD600 between 0.4-0.6. Expression is induced at this point by the addition of IPTG (0.5

mM) and the protein was given an extended, 18 hr, expression time (overnight). The following morning, the cells were harvested as performed for the 4HB mutants. Yields of cell paste (by weight) were similar to those of the 4HB despite the total growth time being more than doubled.

### 2.4.3 Extraction and Purification of 4HB Proteins

A typical purification scheme begins with 1.5 g cell paste. The cells were thawed and resuspended in 20 volumes (30 mL for 1.5 g cell paste) of lysis buffer (10 mM MES pH 6.0, 0.5 M NaCl, 20 mM imidazole) and ruptured by French Press (3 passes) under 15,000 PSI, with ice bath chilling. Care must be taken to ensure a slow, drop-wise, use of the French Press, as haste leads to poor lysis quality. The lysate was then centrifuged for 30 minutes at 22,000 x g and the soluble supernatant decanted and filtered through 0.2 µm membrane syringe-based disc filter (Whatman) prior to chromatography. Analysis of the lysis material (soluble supernatant and insoluble pellet) revealed that only LZEE was soluble upon lysis.

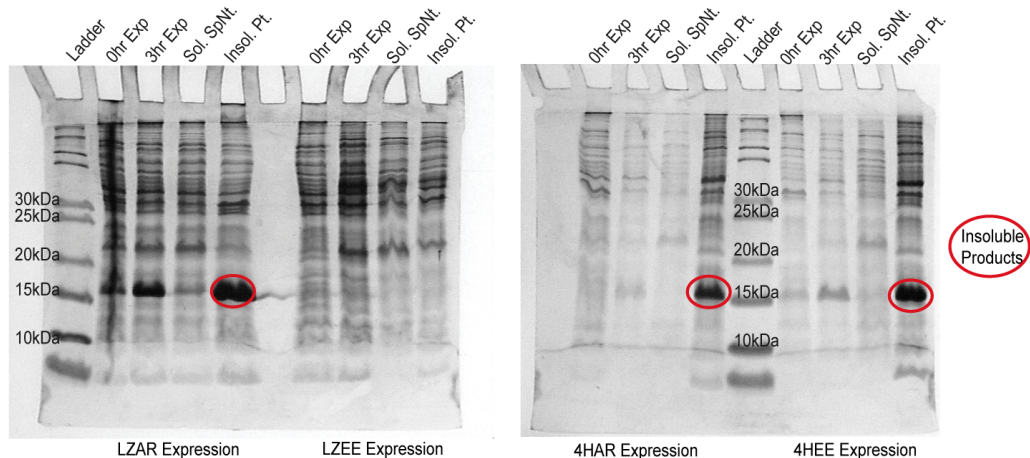


Figure 2.8 Expression of the 4HB constructs. Each expression sample is shown at pre-induction (0 hr) and after 3 hr induction with IPTG [0.5 mM]. The samples were lysed and centrifuged to separate the soluble and insoluble proteins as indicated



Figure 2.8 shows the expression and insoluble nature of 4HAR, 4HEE, and LZAR, which all appeared in the insoluble pellet. LZEE expression could not be confirmed in the experiment shown in Figure 2.8 (left gel). Expression was subsequently confirmed during analysis of the chromatographic purification (see Figure 2.9). It is possible that the low level of expression of LZEE improved solubility or that the expression of a DNA binding protein inhibited expression. Work done by others to express LZEE cloned into pBAD, which uses a regulatable expression system, ultimately failed.

Because of the difference in solubility, the process bifurcated such that LZEE was processed by chromatography directly from the lysis buffer, while LZAR, 4HEE, and 4HAR were processed in 6 M guanidine. For a detailed protocol of the one step purification run using the AKTA FPLC and representative chromatograms, see Appendix 2. The buffer composition for the LZEE chromatography run was as follows:

Equilibration – 10 mM MES pH 6.0, 0.5 M NaCl, 20 mM imidazole.

Elution – 10 mM MES pH 6.0, 0.5 M NaCl, 0.4 M imidazole.

The three insoluble pellets were resuspended in the following equilibration buffer containing 6 M Guanidine and then centrifuged again at 22,000 x g for an additional 30 minutes. The buffer composition for the purification protocol was as follows:

Equilibration – 10 mM MES pH 6.0, 6 M Guanidine, 0.5 M NaCl, 20 mM imidazole.

Elution – 10 mM MES pH 6.0, 6 M Guanidine, 0.5 M NaCl, 0.4 M imidazole.

Purification of each protein was achieved using single step affinity chromatography using  $\text{Co}^{2+}$  charged HiTrap Chelating 1 mL column (GE Healthcare). Cobalt Acetate [0.2 M] was the typical solution used to charge the column. Following the load application to the column, a 5 column volume (CV) wash was performed using Equilibration buffer. The protein was eluted by a 25 CV Equilibration-Elution Buffer gradient from 20 mM to 400 mM Imidazole. The eluent was collected in 1 mL fractions which were analyzed by SDS PAGE on 20 % acrylamide gels (75:1 acrylamide to bis-acrylamide ratio) run using a tricine/SDS buffering system for 1 hr at 120 V. For a representative chromatogram from the HiTrap  $\text{Co}^{2+}$  Chelating step see appendix 2. The typical elution fractions (C1-C5) have been highlighted in the chromatogram, though SDS PAGE analysis is always required to confirm purity.

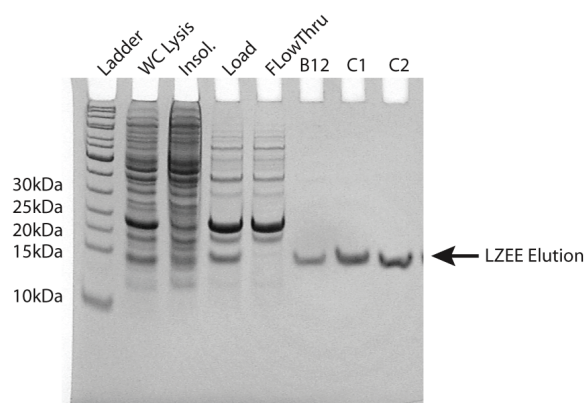


Figure 2.9 Analysis of LZEE purification steps using HiTrap chelating affinity column purification with  $\text{Co}^{2+}$  metal. The gel analysis shows the whole cell lysis (WC), the insoluble pellet, the soluble load, the flow through, and then selected elution fractions B12-C2, corresponding to approximately [240-260 mM] imidazole, showing the target LZEE protein. For the chromatogram for this run, see appendix 2.

Following the chromatography run, the columns could be regenerated for future use by stripping the divalent metal with 0.05 M EDTA followed by a 1 hr 0.5 M NaOH cleaning step, water rinse and reapplication of the  $\text{Co}^{2+}$  solution. Typically purification performance began to noticeably diminish after only 5 runs, although the manufacturer attests to far more column cycles. Due to the relatively low cost of these columns, discarding after 5 runs is the wiser option.

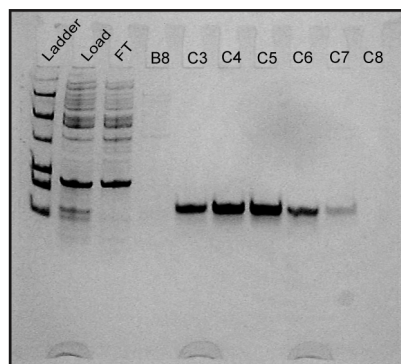


Figure 2.10 Analysis of LZD 73 purification steps using HiTrap chelating affinity column purification with  $\text{Co}^{2+}$  metal. The gel analysis shows the load, the flow through (FT) and selected elution fractions B8 and C3-C8, corresponding to approximately [240-270 mM] imidazole.

#### 2.4.4 Concentration and Buffer Exchange into Storage Buffer

For the soluble proteins (LZEE, LZD73, LZD87, and reverseGCN4), peak fractions were pooled into 4 or 5 mL batches depending on SDS PAGE analysis. The 4-5 mL batches were concentrated to approximately 100-200  $\mu\text{L}$  using a Centricon Ultra 4000 (Millipore), by centrifuging for 30-60 minutes at 10,000  $\times g$  depending on the batch. Peptides purified prior to August, 2010 were buffer exchanged and stored in 10 mM Tris HCl pH 7.7, 150 mM NaCl, and 10 % Glycerol or else 50 mM potassium acetate pH 5.4, 150 mM NaCl, 10 % Glycerol using a Biospin 6 column.

Batches after August, 2010 were buffer exchanged into a storage buffer of 20 mM HEPES pH 7.0, 150 mM NaCl, 10 % Glycerol using a Biospin 6 column that had been prepared by flushing with 4 X 500  $\mu$ L flushes of storage buffer. The switch to HEPES buffer in August, 2010 was made because Tris buffers are known to inhibit potential cross-linking reactions and the acetate based buffer yielded proteins with lower binding activity. Samples were stored at -80 °C where they were shown to remain active for at least 24 months. Samples that were subjected to freeze/thaw cycles saw an immediate decrease in activity and were virtually inactive after three cycles.

Insoluble proteins were stored in elution buffer containing 6 M Guanidine. Attempts to buffer exchange these proteins into less chaotropic buffers such as 0.5 M arginine resulted in precipitation. Prior to their use in EMSA experiments, they were diluted to working concentration in 1X binding buffer. While this method carried the risk of leading to a precipitation, it was hoped that at very low concentrations [ $\leq 100$  nM], the protein would fold into a soluble state without aggregation.

Protein concentrations were primarily determined by UV260 absorption using the estimated extinction coefficient of  $8480 \text{ M}^{-1}\text{cm}^{-1}$  as calculated using ExPASy Protein Parameter analysis of the peptide sequences. Concentrations were also confirmed by gel analysis using SYPRO Ruby stain, where protein samples were fit to a BSA standard curve, using spot quantitation on the Storm Imager (Molecular Dynamics) to measure peptide quantity.

#### 2.4.5 Circular Dichroism Analysis of LZD73

The  $\alpha$ -helical content of LZD73 peptide was analyzed by circular dichroism (CD) spectroscopy. The LZD73 sample was buffer exchanged into a buffer optimized for Enteropeptidase digestion (20 mM Tris-HCl (pH 8.0 @ 25°C), 50 mM NaCl, 2 mM  $\text{CaCl}_2$ ) in anticipation that the N-terminus of the protein might need to be removed. However, the CD was done on uncleaved protein. Analysis of the LZD73 protein at [1.2  $\mu\text{M}$ ], in the absence of DNA, using the buffer alone as background, is shown in Figure 2.11A. For the protein plus DNA sample analysis, a second prep was prepared with [1.2  $\mu\text{M}$ ] LZD73 and [1  $\mu\text{M}$ ] 58mer DNA containing one CREB site (5'-ATGACGTCAT-3') in the middle of the sequence. [1  $\mu\text{M}$ ] CREB site DNA in Enteropeptidase buffer was used as the background measurement for this analysis. The CD analysis for this sample is shown in Figure 2.11B. The 58mer DNA sequence is provided in Appendix 1 and is also depicted in Figure 3.4.

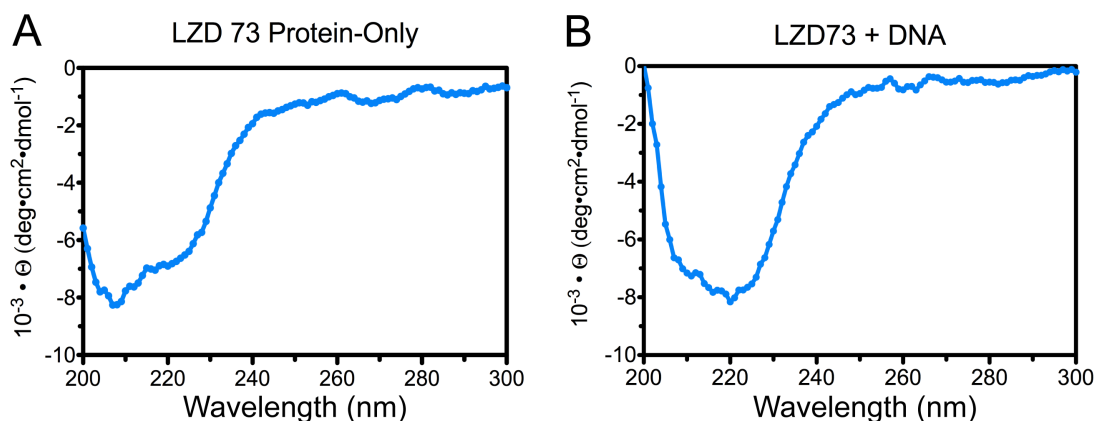


Figure 2.11 CD analysis of LZD73 protein. (A) LZD73 protein alone [1.2  $\mu\text{M}$ ] in enteropeptidase buffer (see above) and (B) LZD73 [1.2  $\mu\text{M}$ ] with 58mer CREB DNA [1  $\mu\text{M}$ ]. The two dips at 210 nm and 220 nm are indicative of  $\alpha$ -helical folding. An increase in the 220 nm signal is observed upon DNA addition, indicating additional folding as the protein binds to the CREB DNA.

The signal produced is typical of  $\alpha$ -helices, indicating this protein exists primarily in that folded form. This is in agreement with our design and can be viewed, in combination with the highly soluble nature of LZD73 and LZD87, as a positive indication that the proteins exist in a well-folded state even in the absence of DNA. Additionally, it is observed that the signal at 220 nM increases when CREB DNA is added to the sample. This indicates that there is additional folding into  $\alpha$ -helices upon binding to DNA, which is in concurrence with prior analysis of other bZip peptides. This data does not indicate whether the protein exists as a dimer when free in solution (not bound to DNA), but previous results on leucine zipper peptides suggested that the monomers are predominantly unfolded. This data strongly supports the notion that our design has led to a protein that folds correctly. The next step of characterizing DNA binding is covered in Chapter 3.

## Chapter 3: Binding Characterization of DNA Looping Proteins by EMSA

### *3.1 Overview*

Electrophoretic Gel Mobility Shift Assays (EMSAs) were used to provide an initial characterization of the protein:DNA interactions. In order for a looping protein to effectively function, it must be able to stably bind two separate DNA fragments simultaneously in a “sandwich complex”. To assess this potential we designed an EMSA assay that would form an “asymmetric” sandwich, meaning the two bound DNA fragments would have significantly different lengths. For comparison, a known single binding peptide consisting of the bZip domain from GCN4 was used as a control (generously provided by Jim Maher from his neutral wild type control ‘PAA’ protein used in electrostatic DNA bending experiments) (McDonald et al., 2007).



### 3.2 *Materials and Methods*

The Electrophoretic Mobility Shift Assay (EMSA), also known as a gel shift assay, is a common technique for characterizing protein:DNA interaction. DNA migrating through a polyacrylamide, driven by an electric field, will migrate slower if protein is bound to it. As a proof of concept for our protein design, we required an assay to show that our DNA looping proteins possess an ability to bind two independent strands of DNA in a stable structure termed a “sandwich complex”. Any looping protein will, by definition, be required to bind DNA in this manner. This chapter describes the EMSA experiments performed to evaluate whether the 4HB and LZD proteins can bind two separated helices of DNA and form a stable sandwich complex.

#### 3.2.1 Binding/Ligation Buffer Formulation

The confirmation of a sandwich complex will constitute a successful first round of characterization. Further analysis to demonstrate looping will involve a series of ligation dependent experiments that necessitates a buffer compatibility between EMSA binding experiments and T4 DNA ligase activity. The NEB provided buffer for T4 DNA ligase consists of 50 mM Tris HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, 1 mM ATP. This formulation was an initial source of difficulty. The GCN4 bZip peptide would not shift in EMSA studies when the MgCl<sub>2</sub> concentration was at 10 mM, and T4 DNA ligase would not function at our initial binding buffer concentration of 2 mM MgCl<sub>2</sub>. Additionally, our initial binding buffer contained a high concentration of BSA (2 mg/mL) meant to protect the leucine rich peptides or

the ligase enzyme itself from adsorbing onto the walls of the polypropylene reaction tubes, but this combination of BSA, MgCl<sub>2</sub> and DTT led to a rapid precipitation.

Optimization by trial and error was done to identify a buffer formulation that would allow DNA binding and ligation to occur in a precipitate free solution. A successful binding/ligation (B/L) buffer formula was achieved using 50 mM Tris HCl pH 7.7, 4 mM NaCl, 4 mM KCl, 4 mM MgCl<sub>2</sub>, 2 mM ATP, 0.2 % Glycerol, 100 µg/mL BSA, 10 mM DTT, and 0.01 % NP40 (IGEPAL).

### 3.2.2 Sample Preparation and Gel Analysis

All dilutions were done using 1X B/L buffer. For EMSA assays, protein and DNA were added in equal volume (5 µL each), mixed by pipette, and incubated for 10 minutes at room temperature. 2 µL of 6X DNA loading dye was added, and the sample was immediately loaded onto the gel.

Three EMSA gel formulations were utilized over the course of these experiments:

10 % acrylamide (75:1 acrylamide:bis-acrylamide) was used with a TBE buffer (50 mM Tris HCl pH 8.1, 50 mM boric acid, 1 mM EDTA) on the 25mer EMSA with the 4HB (Figure 3.1). Binding buffer (used prior to the Binding/Ligation buffer formulation) was composed of 10 mM Tris HCl pH 7.8, 1 mM EDTA, 5 mM KCl, 5 mM MgCl<sub>2</sub>, 0.01 % NP40, 50 µg/mL BSA). The gel was run for 1 hour at 400 volts.

7 % acrylamide (75:1) was cast with TBE buffer (50 mM Tris HCl pH 8.1, 50 mM boric Acid, 1 mM EDTA), which was also used as the running buffer, for the 145 bp and 177 bp EMSA with GCN4 bZip, LZEE, and LZD73. By this point it had been

decided that this binding buffer should be compatible with T4 DNA ligase (ATP dependent) so the buffer had been reformulated to the B/L described above. The gel was run for 1 hour at 400 volts.

To resolve the EMSA sandwich complexes sufficiently, an acrylamide percentage lower than 1 % (75:1) was required. Below 6 % at 75:1 acrylamide:bis-acrylamide, the polymerized product begins to exist in a liquid state. To get around this problem required the invention of a new hybrid gel approach that combined 0.5 % agarose with either 4 % or 5 % acrylamide. The formulation used in the EMSA gels in Figures 3.5, 3.6, and 3.7, the 58mer/30mer sandwich complexes, is as follows: 5 % acrylamide (75:1), 0.5 % agarose, run in 1X TBE Buffer (50 mM Tris HCl pH 8.1, 50 mM Boric Acid, 1 mM EDTA). Gels were run for 1 hr at 300 volts.

The gels were dried on filter paper and were used to expose storage phosphor screens overnight. The images were captured using the Storm 860 Phosphorimager (Molecular Dynamics) scanner and visualized using ImageJ.

### *3.3 Results*

#### *3.3.1 EMSA Analysis of 4HB mutants*

The migration of a sandwich complex in an acrylamide gel should differ greatly from that of a single DNA fragment with a singly bound protein. By comparing the control GCN4 bZip binding behavior to the 4HB proteins, it should be evident whether the designed proteins are binding DNA according to design. GCN4 bzip has been previously shown to bind with high affinity to both the palindromic

CREB and the pseudo-palindromic AP1 site DNA (Koldin, Suckow, Seydel, Wilcken-Bergmann, & Müller-Hill, 1995). For the initial 4HB construct EMSA experiments, 23 bp DNA fragments containing the AP1 binding site (5'-ATGACTCAT-3') were end-labeled with  $^{32}\text{P}$  using polynucleotide kinase (PNK) with  $\gamma\text{-}^{32}\text{P}\text{-ATP}$  and held at a constant concentration of 5 nM. Each of the four peptides was then added in at 5 nM, 25 nM, and 100 nM monomer concentrations (see Figure 3.1). Assuming a tetrameric folding of the peptide, the molar ratios of protein to DNA would therefore be 0.25:1, 1.25:1, and 5:1, respectively. The results demonstrate that both 4HAR and LZAR, two of the three insoluble mutants, aggregate along with the DNA and result in a well shift. The intensity of this well

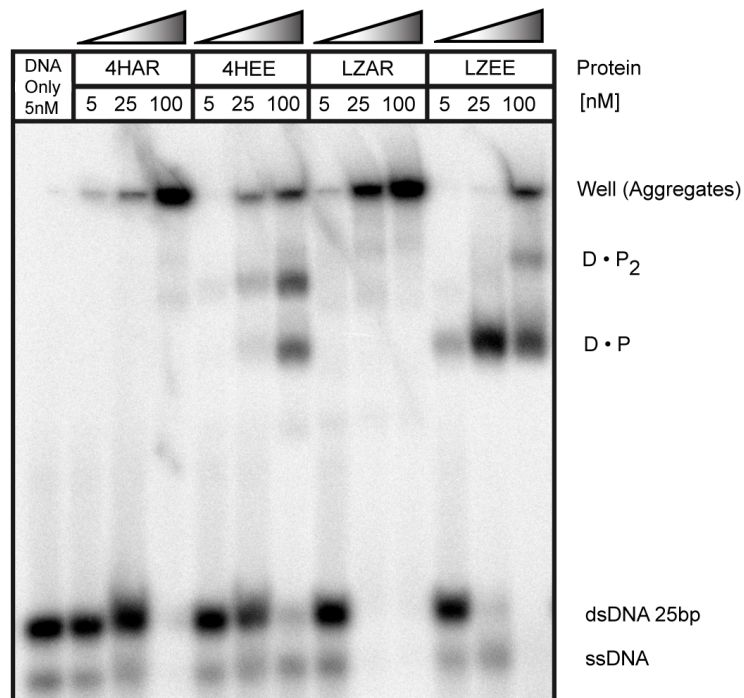


Figure 3.1 EMSA of the 4HB construct with 23 bp DNA (5'-AGTGGAGATGACT-CATCTCGTGC). DNA concentration was held constant at [5 nM] in all lanes while the 4 proteins (4HAR, 4HEE, LZAR, and LZEE) were added in a concentration gradient of 5, 25, and 100 nM. The well shifts observed are likely a result of insoluble aggregates. The complexes that stably migrated through the gel have been indicated

aggregate is proportional to the protein concentration and there is no distinguishable population that migrates into the gel. This suggests that 4HAR and LZAR are not folded into a stable tetramer and that their insolubility is not enhanced by the low concentration. Work with these two mutants was not continued.

The 4HEE and LZEE mutants however did form stable DNA-bound complexes that migrated as distinct populations in the gel shift assay. The two populations observed in the Figure 3.1 could either be sandwich complex or perhaps singly bound 23mer and doubly bound 23mer. In order to make sense of the binding, the GCN4 bZip single binding control protein was used to further characterize the binding pattern. Figure 3.2 illustrates the binding of GCN4 bZip peptide of increasing concentration with a 145 bp DNA fragments containing one CREB site.

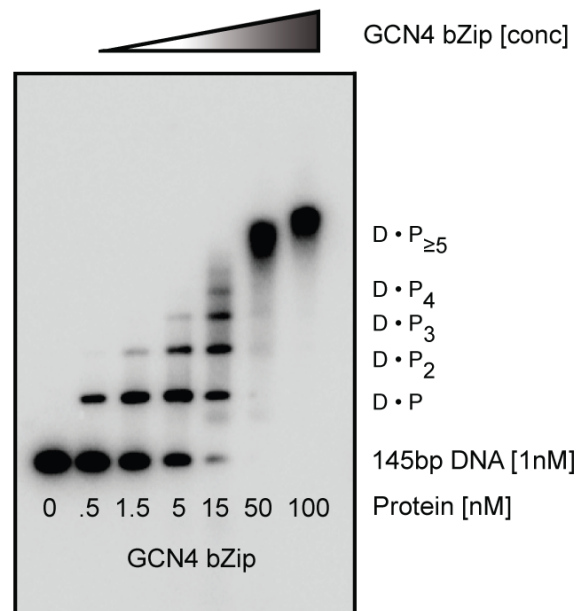


Figure 3.2 EMSA of 145 bp DNA with GCN4 bZip single binding protein control. The 145 bp DNA with single CREB site DNA (5'-ATGACGTCAT-3') was held constant at 1 nM while the GCN4 bZip single binding site control peptide was titrated in at increasing concentrations

This gel illustrates the sequential binding events as the peptide concentration is increased. This DNA fragment only contained one CREB binding site, meaning additional binding must be through non-specific interactions. This is likely due to the low ionic strength of the binding buffer (4 mM NaCl, 4 mM KCl, 4 mM MgCl<sub>2</sub>). This formulation was similar to that used in prior binding characterization studies, however those also contained 200 mM guanidinium and 10 % glycerol (Chan, Fedorova, & Shin, 2007; McDonald et al., 2007). The ionic strength of this buffer is therefore much higher due to Guanidine HCl salt existing entirely as the guanidinium cation (pKa 13.6) in solution. Mimicking this buffering component would not be feasible because of the previously mentioned compatibility requirement for T4 DNA

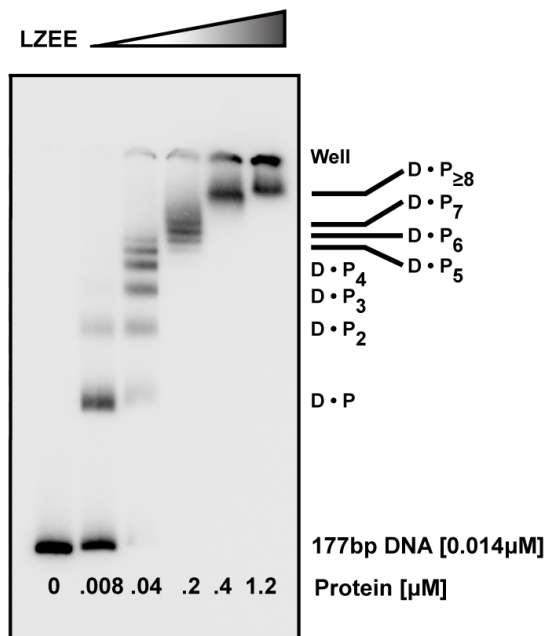


Figure 3.3 EMSA of LZEE peptide and 177 mer DNA. The 177 bp DNA [1 nM] contains both the CREB site (5'-ATGACGTCAT-3') and Inv-2 site sequence (5'GTCATATGAC-3') separated by 115 bp. The DNA was held constant while the LZEE peptide was titrated in at increasing concentrations

ligase. To determine whether LZEE was binding as a sandwich complex or as a singly bound peptide, further analysis with a 177 bp DNA fragment was performed using a similar protein gradient approach.

This gel revealed a similar pattern for binding (see Figure 3.3) as seen in the GCN4 bZip control (Figure 3.2) indicating LZEE did not forming a sandwich complex with DNA. Because 4HEE and LZEE behaved nearly identically in the original 23 bp EMSA, both can be ruled out as potential DNA looping proteins.

### 3.3.2 EMSA Analysis of LZD Proteins

In order to demonstrate the existence of a stably bound sandwich complex an EMSA technique was employed where the CREB and Inv-2 specific site would bind DNA fragments of considerably different lengths (58 bp and 30 bp, respectively).

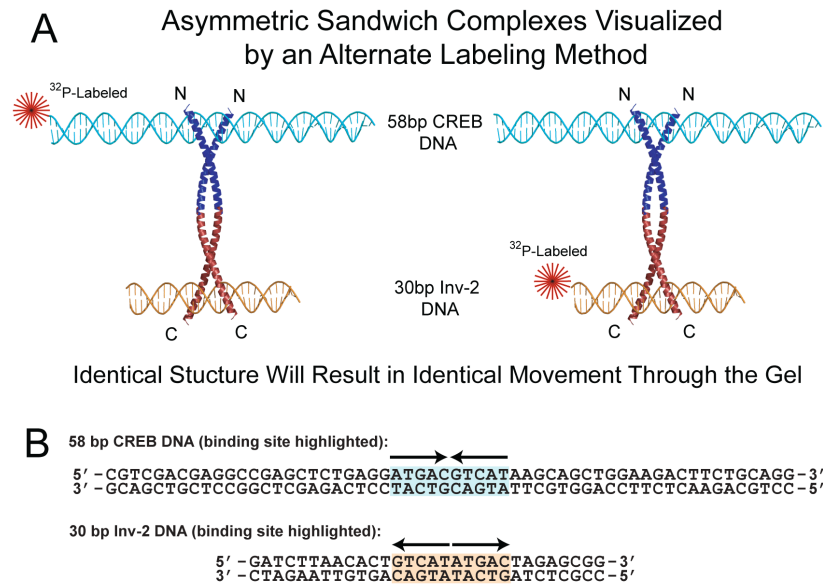


Figure 3.4 (A) Image illustrates the asymmetrical sandwich complex that would confirm dual binding of the LZD protein. By radiolabeling only one of the two double helical DNA's at a time, the resulting mixed population will be uniquely produced in both sets of parallel experiments. (B) 58mer and 30mer DNA with CREB and Inv-2 binding sites. The sequence inversion has been highlighted with arrows.

This EMSA method, termed the asymmetric approach, should unequivocally demonstrate a sandwich population by  $^{32}\text{P}$  labeling one of the fragments while titrating in the unlabeled counterpart. If this was done in parallel – the 58mer is  $^{32}\text{P}$  labeled while the 30mer is not, or vice versa – a unique population should emerge in both sets that can only be explained by the asymmetric sandwich complex. Figure 3.4 illustrates this parallel approach to sandwich formation.

The GCN4 bZip peptide and the *reverseGCN4* peptide were used as single binding controls in this experiment. All three proteins were subjected to identical procedures. The GCN4 bZip EMSA is shown in Figure 3.5, while the *reverseGCN4* is shown in Figure 3.6, and LZD73 in 3.7.

The GCN4 bZip peptide exhibited tight binding to the CREB DNA and weak binding to the Inv-2 site DNA at low concentration of peptide (Figure 3.5 lanes 2 and 7). However, at higher concentration it showed significantly more promiscuous binding than the *reverseGCN4* C-terminal single binding control (see Figure 3.6). As seen in Figure 3.5 lane 4, the 58mer CREB DNA shifts to two additional populations upon the addition of excess protein and with the Inv-2 site DNA shifts to one additional population (lane 9). This non-specific binding can be competed off with addition of the cold competitor and does not lead to a unique population that would be expected if an asymmetric sandwich complex were forming. The non-specific binding is likely a result of the low ionic strength of the buffer system used. This is seen as unavoidable given the requirement that this buffer must serve as both a binding and a ligation buffer.



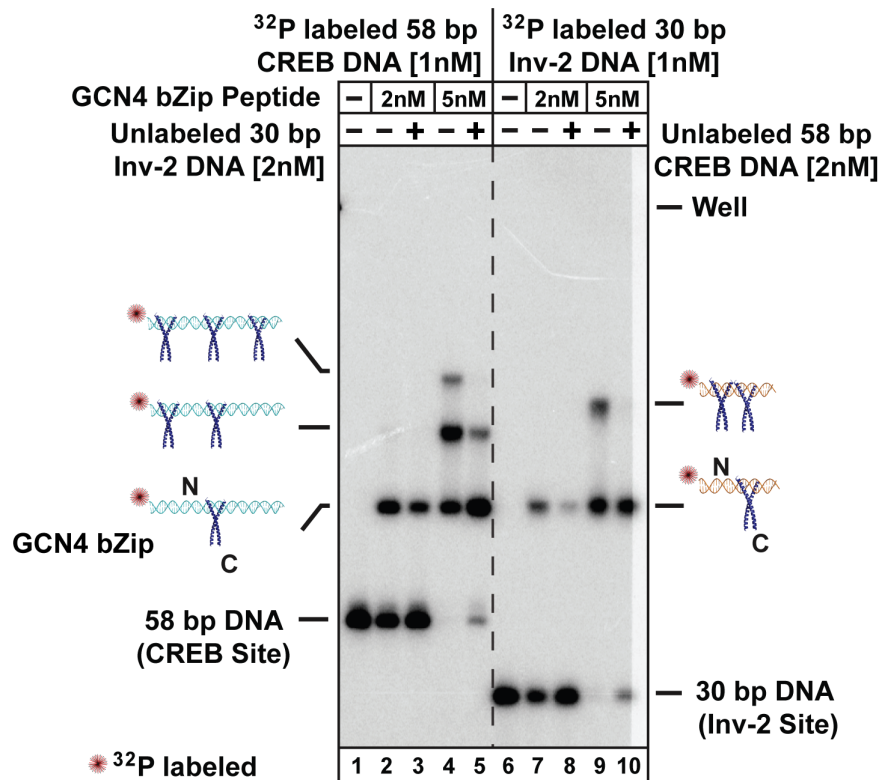


Figure 3.5 EMSA of GCN4 bZip single binding control with DNA fragments of 58 bp (CREB site) or 30 bp (Inv-2 site). The peptide contains one DNA binding region at the N-terminus of the peptide and shares amino-acid identity with the wild-type bZip region of GCN4. The radiolabeled DNA fragments were held constant at [1 nM] (lanes 1-5 CREB 58mer, lanes 6-10 Inv-2 30mer) while the GCN4 bzip protein was added at [2 nM] and [5 nM]. Unlabeled DNA fragments with opposite binding sites were then added at [2 nM] as cold competitors (lanes 3,5,8,10).

The *reverse*GCN4 demonstrates tight binding to the Inv-2 site (lanes 7 and 9) and this binding cannot be competed off by the addition of the CREB site DNA (lanes 8 and 10). The peptide does exhibit binding to the CREB DNA (lanes 2 and 4), though this is of lower affinity than for Inv-2 DNA. This difference in affinity is confirmed by competing CREB DNA off with Inv-2 DNA (lane 3 - completely, and lane - partially), but not vice versa (lanes 8 and 10). This result of binding CREB DNA alone is not a complete surprise because of the similarities between

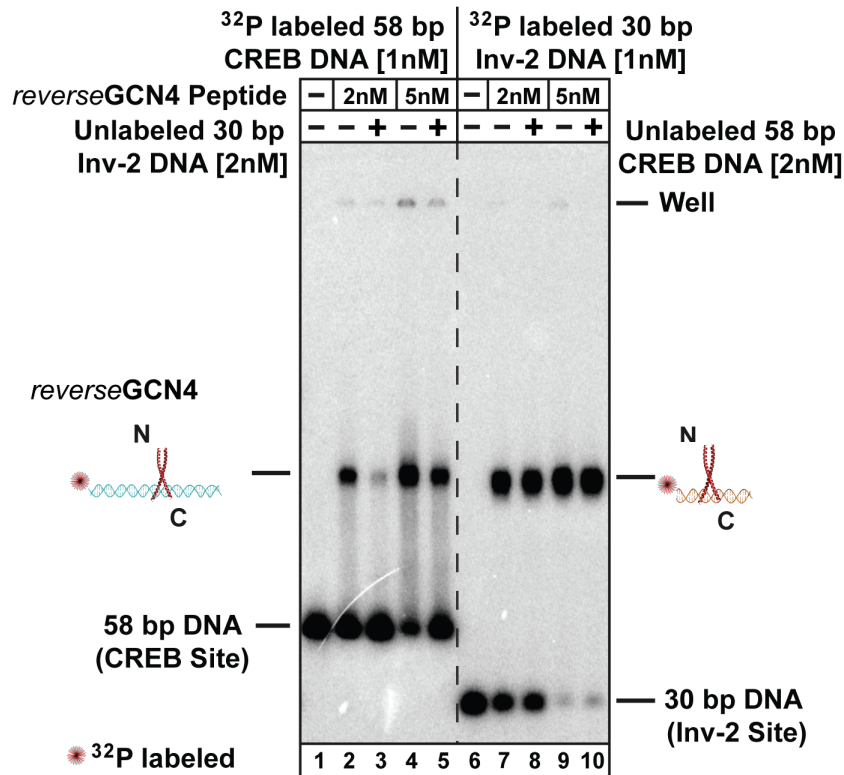


Figure 3.6 EMSA of *reverseGCN4* bZip single binding control with DNA fragments of 58 bp (CREB site) or 30 bp (Inv-2 site). The peptide contains one DNA binding site at C-terminus of the peptide. The radiolabeled DNA fragments were held constant at [1 nM] (lanes 1-5 CREB 58mer, lanes 6-10 Inv-2 30mer) while the GCN4 bzip protein was added at [2 nM] and [5 nM]. Unlabeled DNA fragments with opposite binding sites were then added at [2 nM] as cold competitors (lanes 3,5,8,10).

binding sites. This gel clearly shows single binding of the C-terminal protein binding domain with both CREB and Inv-2 site DNA, with a stronger affinity for the Inv-2 site.

Figure 3.7 shows the EMSA results from LZD73 binding, which proved far more interesting. It should be noted that our predicted asymmetric EMSA product only appears, as expected, in lanes 3, 5, 8, and 10. The band can be seen paired with the cartoon representation that matches the asymmetric graphic from Figure 3.4. The promiscuous binding seen in the GCN4 bZip control gel is again seen in this gel,

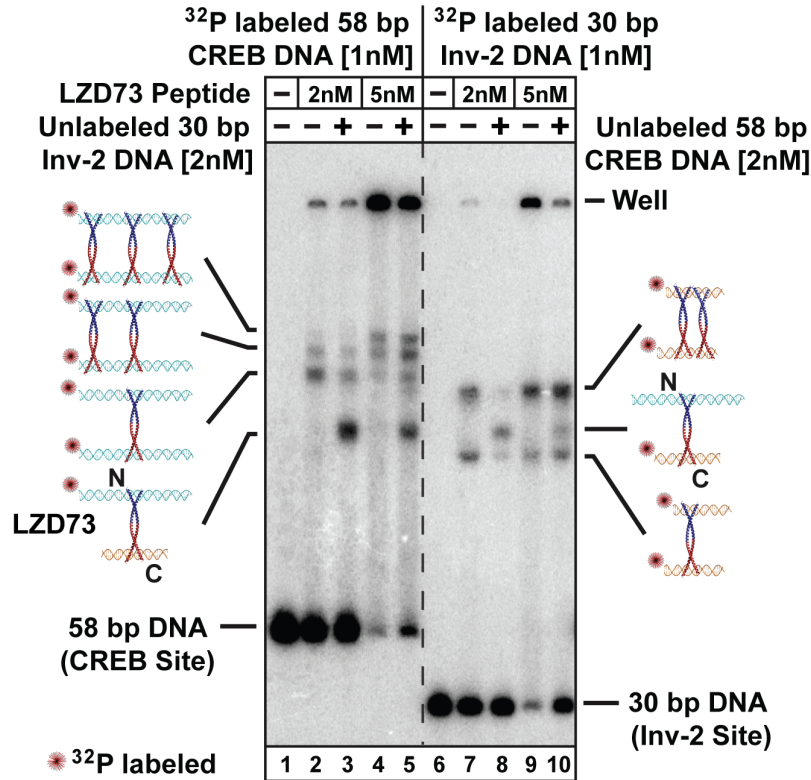


Figure 3.7 EMSA LZD73 with DNA fragments of 58 bp (CREB site) or 30 bp (Inv-2 site). The radiolabeled DNA fragments were held constant at [1 nM] (lanes 1-5 CREB 58mer, lanes 6-10 Inv-2 30mer) while the GCN4 bzip protein was added at [2 nM] and [5 nM]. Unlabeled DNA fragments with opposite binding sites were then added at [2 nM] as cold competitors (lanes 3,5,8,10). The emergence of the sandwich complex in lanes 3,5,8,and 10 confirms the dual binding capability of the LZD protein

however here it leads to a variety of sandwich complex products that have been interpreted in the figure. What is important about this gel is that the non-specific sandwich complexes (either 58mer-58mer or 30mer-30mer) are converted to the asymmetric sandwich (illustrated in Figure 3.4) as expected upon addition of the unlabeled competitor. The existence of this predicted sandwich complex product is a satisfactory result and establishes a proof of concept for the LZD design.

### 3.4 *Discussion of Results*

#### 3.4.1 The 4HB Mutant Binding

In order for DNA looping to occur, two DNA binding events must take place. The 4HB mutants relied on a tetrameric design where complex folding was required in order to incorporate two DNA binding sites. The EMSA results shown in Figure 3.1 indicate that LZEE and 4HEE do not fold into tetramers with dual binding capacity. The single binding pattern of LZEE (the only soluble mutant) in Figure 3.3 paralleled that of the GCN4 bZip control protein (Figure 3.2), from which we can argue that it folded as a dimer. This is quite easy to imagine since the major domain of the protein is from GCN4's dimeric leucine zipper. Investigations into the nature of coiled-coil folding and oligomerization have identified a trigger sequence for the dimerization of GCN4 (Kammerer et al., 1998). The theory of trigger sequences has been further resolved and recently applied broadly to all bZip proteins and additional states of oligomerization (Ciani et al., 2010; Steinmetz et al., 2007). The dimerization trigger sequence for GCN4 was identified as –YHLENEVARLKKL–. In our 4HB peptide constructs, this comprises the 13 aa immediately prior to the 7 aa linker. Additionally, the 7 aa linker of LZEE –VEELLSK– was taken directly from the leucine zipper region of GCN4 (aa257-aa263 from the wild type protein, or aa68-aa74 in our peptides) and the heptad repeat maintains V and L at the a' and d' position within the zipper. Furthermore, the leucine zipper motif of hydrophobic residues at the a' and d' positions is maintained during the transition to and then throughout the tetramerization domain. Based on the available literature on the

folding of leucine zipper oligomers (cited above), the conclusion that LZEE and 4HEE simply folded into dimeric form is well founded.

While the insoluble nature of 4HAR and LZAR likely meant they simply formed unstructured aggregates, LZEE and 4HEE clearly had defined structure and could bind DNA. While the failure of the 4HB design represented a major setback for this project and resulted in work with these four proteins being suspended, it can be argued that the solubility of LZEE was a small, albeit orthogonal, success in its own right. We demonstrated that the DNA binding capacity of LZEE was not diminished by elongating the wild-type GCN4 leucine zipper. This proves, at the very least, that a bZip dimer can be engineered at the C-terminal and still readily bind DNA at the N-terminal end. This result would prove useful during the redesign process.

### 3.4.2 The LZD Mutant Binding

In order for DNA looping protein to function, it must be capable of tightly binding two distinct helices of DNA. LZD73 demonstrates this feat as seen in Figure 3.7. Of concern is the promiscuous binding of the N-terminal GCN4 binding domain. This binding behavior can likely be attributed to the low ionic conditions in the binding buffer (4 mM NaCl, 4 mM KCl is well below standard cellular conditions). In order to provide a biochemical analysis of DNA looping with the LZD proteins, it was decided that ligase-mediated ring-closure experiments would be utilized. This forward looking plan required the binding buffer to serve in both EMSA assays and ligation reactions with protein binding, and the resulting low salt was a necessary compromise. The NEB provided T4 DNA ligase buffer contains 50 mM Tris HCl, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 1 mM ATP, pH 7.5. The absence of monovalent cations is

not accidental. During the experiments that sought to reconcile the buffer conditions it was observed that ligase activity drastically decreased above 20 mM NaCl. T4 DNA ligase activity is essential to the experiments in chapter 4 and 5, so it was decided that the 4 mM NaCl and 4 mM KCl formulation was acceptable. Care would be taken to control the protein:DNA ratios and concentrations would be kept as low as practically possible to mitigate potential non-specific interactions.

## Chapter 4: Length Dependent Loop Formation Using Ligase-Mediated DNA Dimerization

## 4.1 Overview

In forming, DNA loops must overcome two free energy hurdles. The first is entropic, where the likelihood of two ends of DNA meeting in a loop decreases as the distance between them increases. This factor is independent of the geometry of DNA and as such is in contrast with the second factor, the torsional rigidity of DNA double helix. For large rings >2000 bp, the rigidity plays virtually no part in loop but as the DNA decreases in length, this factor becomes considerable (Horowitz & Wang, 1984). The LZD proteins were designed to investigate this property of DNA as it pertains to protein-mediated looping. The characterization of LZD found that the two ends could independently bind to DNA. In principle, if such dual binding is possible, then looping must also be possible providing the length of DNA in the loop is

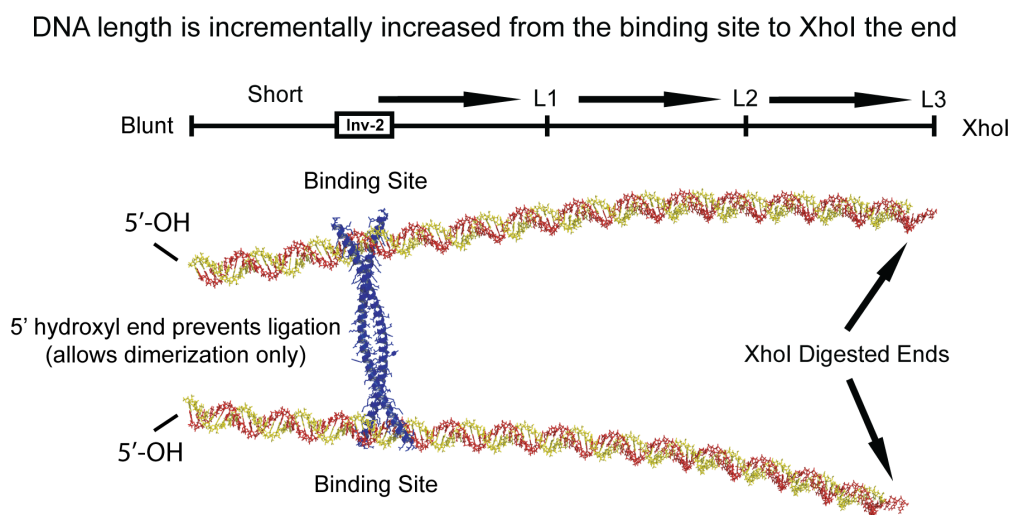


Figure 4.1 Overview of dimerization bridged by sandwich complex formation (lengths not to scale). This illustrates the concept of using sandwich complex formation to increase the likelihood of forming dimer products. At the moment of ligation, a transient loop is formed if both binding sites remain bound to the protein.



sufficiently long that the torsional rigidity is not too high an energetic obstacle. In order to address whether looping could occur, we used a ligation based technique where DNA bound in sandwich complex with protein demonstrates enhanced dimerization versus diffusion controlled DNA-only. Figure 4.1 illustrates the concept behind these experiments.

This approach has been previously utilized to show that c-Myc can form a weakly stable tetramer (Ferré-D'Amaré et al., 1994). This method is difficult to interpret quantitatively because of the multitude of sandwich complexes that can form. But this approach does allow for the qualitative assessment of whether a loop formed, since at the moment of ligation (if it is sandwich complex facilitated) a transient loop must occur. The confirmation of a transient loop was very useful in defining what length ranges to explore with more quantitative methods.

## *4.2 Materials and Methods*

All reagents were purchased from Fisher Scientific with the exception of [ $\alpha^{32}\text{P}$ ]dATP, which was purchased from either Perkin Elmer or MP Biomedicals, and the dNTPs, which were purchased from NEB. All enzymes were purchased from New England Biolabs (NEB). Qiaquick PCR Cleanup kits were purchased from Qiagen.

### *4.2.1 PCR Generation of Variable Length Fragments.*

The variable length DNA fragments were constructed using PCR from a plasmid template (pIx Dimer – previously referred to as pVx6 II) containing the Inv-2 binding site (5'-GTCATATGAC-3'). A constant forward primer was used while the

reverse primers were used to adjust the length of the fragments. The reverse primers also contained an XhoI restriction site at a region 5' to the region that would anneal to the template, while the forward primer, provided by IDT, lacks a 5' Phosphate, which prevents ligation of that end. Figure 4.2 illustrates how the plasmid template was used to produce the variable length fragments (A) and lists the primers used in the experiment (B). The complete sequences of the 3 PCR products can be found

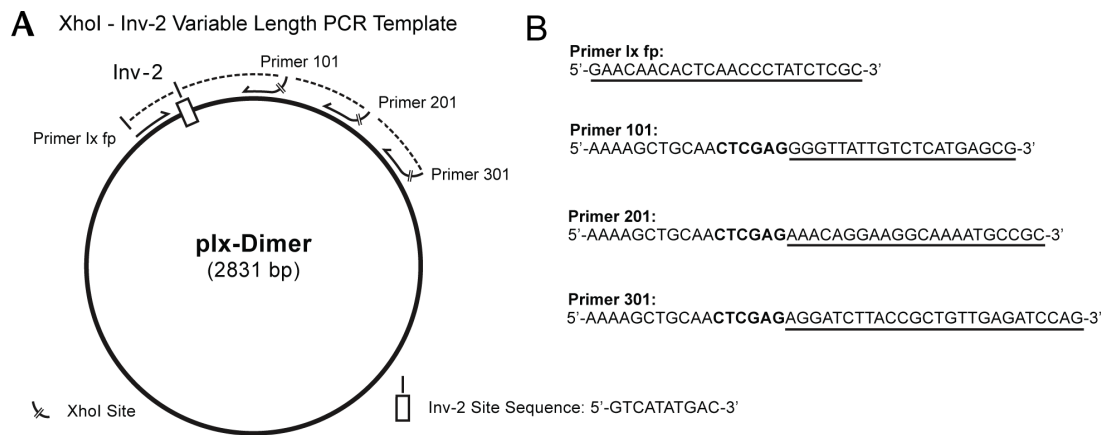


Figure 4.2 Plasmid diagram (A) and primer usage (B) illustrating the construction of the three Ix DNA fragments with Inv-2 binding sites. Underlined primer regions indicate the portion of which anneals to the template while the XhoI sites (CTCGAG) are in **BOLD**.

in Appendix 1. The DNA was body radiolabeled during synthesis by incorporating [ $\alpha$ - $^{32}$ P]dATP into the reaction. Based on the molar ratio of [ $\alpha$ - $^{32}$ P]dATP to unlabeled dATP and the number of dA residues per DNA fragment, the concentrations for the final products could be calculated using scintillation counting. This method of direct measurement, as opposed to using scintillation fluid, is referred to as Cerenkov counting (Plesums & Bunch, 1971).

A PCR master mix of all reagents except primers was prepared as follows: 45  $\mu$ L 5X Phusion buffer, 12.25  $\mu$ L dATP [2 mM], 12.25  $\mu$ L dTTP [2 mM], 12.25  $\mu$ L dGTP [2 mM], 12.25  $\mu$ L dCTP [2 mM], 5  $\mu$ L pVx6 II template (10 ng/uL stock), 80

$\mu\text{L H}_2\text{O}$ , 9  $\mu\text{L}$  [ $\alpha\text{-}^{32}\text{P}$ ]dATP [3.33  $\mu\text{M}$ ], 2.5  $\mu\text{L}$  Phusion polymerase (2 U/ $\mu\text{L}$ ). At running conditions, the dNTP's were [100  $\mu\text{M}$ ] each and the [ $\alpha\text{-}^{32}\text{P}$ ]dATP was [0.1332  $\mu\text{M}$ ] for a ratio of 750:1. Primers were diluted to [2  $\mu\text{M}$ ] and 5  $\mu\text{L}$  of forward primer was added to each tube and 5  $\mu\text{L}$  of the corresponding reverse primer was added to the appropriate tube along with 40  $\mu\text{L}$  of the master mix (50  $\mu\text{L}$  total), for final primer concentrations of [200 nM] each. The reaction protocol was as follows:

Initial denaturation: 99 °C for 3 min

Cycle (33 repeats)

1. 95 °C for 30 sec
2. 59 °C for 20 sec
3. 73 °C for 20 sec

After PCR, but prior to XhoI digestion, the samples were subjected to a Qiaquick PCR cleanup column to remove polymerase and unincorporated nucleotides, and eluted in 50  $\mu\text{L}$  of  $\text{H}_2\text{O}$ . This step was essential for producing a high yield of ligatable DNA fragments with functional XhoI 5' overhangs. The samples were digested with 20 units XhoI for 1 hr in 57  $\mu\text{L}$  of NEB Buffer 2 (10 mM Tris-HCl, 50 mM NaCl, 10 mM  $\text{MgCl}_2$ , 1 mM dithiothreitol, pH 7.9) at 37 °C. The digested products were purified by PAGE (7 % acrylamide, 75:1) run in TBE for 1 hr at 400 V, excised from the gel and then eluted overnight in 500  $\mu\text{L}$  of 50 mM potassium acetate, 1 mM EDTA, pH 7.1. The eluted samples were concentrated to 100  $\mu\text{L}$  by speedvac and then subjected to a second Qiaquick PCR cleanup column. The samples were eluted in water and quantitated using Cerenkov counting. UV260 absorbance was not used to calculate concentration because the final volume of the

eluted products, 50  $\mu$ L, would require 2X dilution to fill the cuvette and the amount of DNA at that volume would not lead to an accurate reading given the large background noise introduced by residuals from the gel extraction process.

#### 4.2.2 Ligation Procedures

DNA samples were diluted to 5 nM in B/L buffer. Protein samples were diluted to 125 nM in B/L Buffer. The proteins used in the experiment were the GCN4 bZip peptide, the *reverse*GCN4 peptide, LZD73 and LZD87. A DNA-only control was also performed. 22  $\mu$ L of DNA [2 nM] and 22  $\mu$ L protein [125  $\mu$ M] were mixed (1X B/L was used in place of protein for the DNA-only sample) and were allowed to equilibrate for 10 minutes at room temperature. 4  $\mu$ L was removed for the time 0 sample then 10  $\mu$ L of T4 DNA ligase (NEB) was added (15 U/ $\mu$ L in B/L buffer). The final concentrations during ligation were 2 nM DNA, 50 nM Protein (a 12.5:1 protein(dimer):DNA ratio), and T4 DNA ligase at 3 U/ $\mu$ L. The reaction was allowed to proceed at room temperature. At time points 2.5, 5, 10, 30, 60, 150, 300, and 1080 minutes, 5  $\mu$ L of the ligation mix was removed and the reaction quenched by adding 3  $\mu$ L of 2 mg/mL Proteinase K in B/L buffer that was brought to 50 mM EDTA. The high concentration of EDTA in the Proteinase K mix should chelate the  $Mg^{2+}$  thus inactivating T4 DNA ligase very rapidly.

The quenched samples were incubated at 37 °C for 15 minutes to allow Protease K to digest the added proteins and ligase. The samples were then stored at 80 °C until reaction was complete (the 1080 minute samples reacted overnight).

The samples were resolved on a 6 % acrylamide gel (75:1) in TBE buffer with 7.5 mM NaCl added (this was shown to improve band resolution). 2  $\mu$ L of 6X DNA

loading dye (30 % glycerol, 0.25% bromophenol blue, 0.25 % xylene cyanol, in H<sub>2</sub>O) was added, the samples mixed and then loaded onto the gel. The gel was run at 100 V (5 V/cm) for 4 hr. The gels were dried and exposed to a storage phosphor screen. The images were acquired using the Storm Phosphorimager and the bands were quantitated using the volume integration function in ImageQuant (Molecular Dynamics). The fraction of dimer product formed was calculated by dividing the dimer population by the sum of the dimer and the monomer population. This value was plotted as a function of time using Prism 5.

### 4.3 Results

The three DNA lengths used would lead to loops with lengths of 208 bp (Ix1), 314 bp (Ix2), and 610 bp (Ix3). The sandwich complex formation holds two DNA molecules in close proximity in space and it may also align the XhoI ends to each other. Both of these effects will lead to enhanced kinetics of ligation. Therefore and increase in the initial ligation rate compared to the DNA-only sample (or single binding peptide controls GCN4 bZip or *reverse*GCN4) can only be attributed to protein induced loop formation or network formation.

The protein and DNA concentration used were chosen based on initial runs that were optimized by trial and error. DNA concentrations that were too low would not yield much if any product because of second order nature of the reaction. At excess protein concentrations, sandwich complexes will not form. On the other hand, protein concentrations that are too low (a low protein:DNA ratio) do not produce

noticeably different results from DNA-only controls, because most of the DNA is not bound.

The gels showing the dimerization of Ix1, Ix2, and Ix3 are presented in Figures 4.3, 4.4, and 4.5 respectively. The Complete data set for the dimer product fraction versus time in plotted in Figure 4.6. This graph shows the complete data set. The initial dimerization rates for LZD73 (blue) and LZD87 (purple) are clearly greater than for DNA-only (red) and the single binding protein controls reverseGCN4 (green) and GCN4 bZip (teal). This represents a successful result and suggests that both LZD73 and LZD87 can form DNA loops as short as 208 bp.

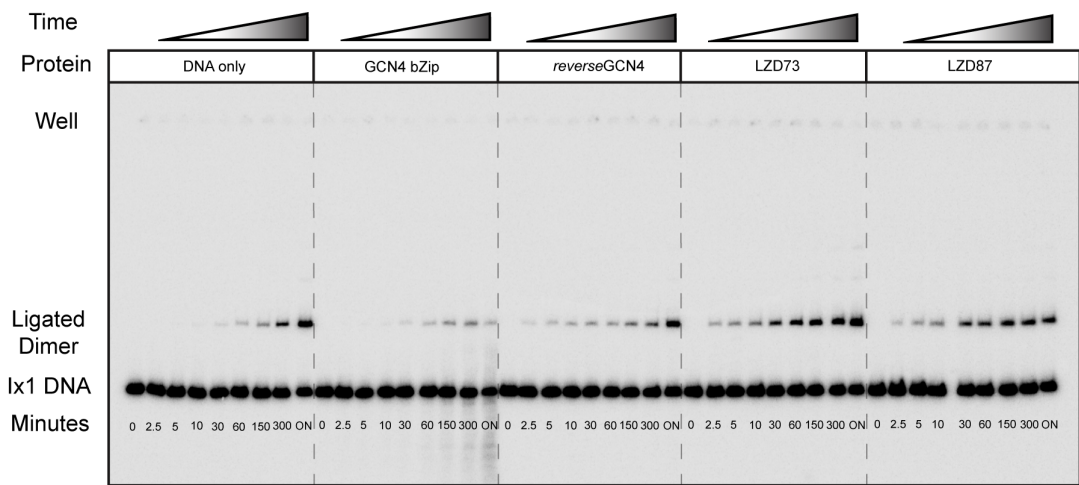


Figure 4.3 Dimerization of Ix1 DNA (29 bp blunt end to Inv-2 site, 104 bp Inv-2 site to XhoI end). Five dimerization experiments for 4 protein samples and a DNA-only control are shown. An increase in the rate of formation of dimer product (ligated dimer) indicates an interaction between one protein and two DNA fragments. Time points are given in minutes. The ON (overnight samples) were allowed to react for 1080 minutes

It should be noted that the GCN4 bZip sample shows evidence of DNA degradation. This is likely due to the presence of a contaminating DNA nuclease. Previous work with this protein sample and DNA consisted of gel mobility shifts in which the DNA was incubated with the protein for only ten minutes prior to loading

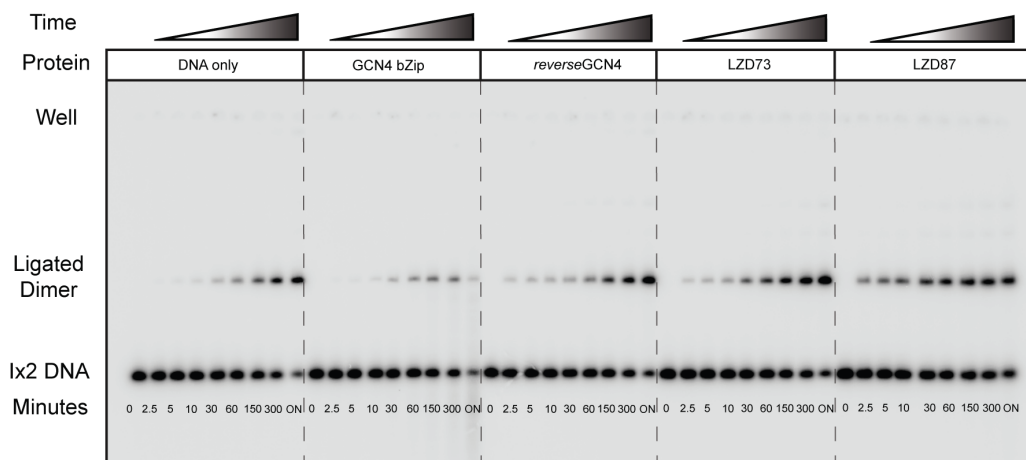


Figure 4.4 Dimerization of Ix2 DNA (29 bp blunt end to Inv-2 site, 207 bp Inv-2 site to XhoI end). Five dimerization experiments for 4 protein samples and a DNA-only control are shown. All conditions and interpretations are the same as for Ix1 ligation.

onto a gel. The gel then acted to separate the DNA from the contaminant and therefore any degradation was minimal and not recognized. The extended incubation time for this experiment (the ON –overnight- samples were incubated for 18 hours)

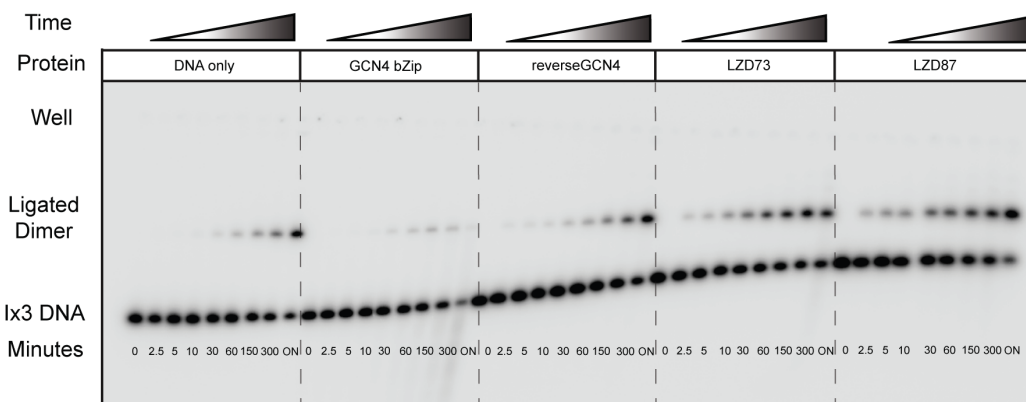


Figure 4.5 Dimerization of Ix3 DNA (29 bp blunt end to Inv-2 site, 305 bp Inv-2 site to XhoI end). Five dimerization experiments for 4 protein samples and DNA-only control are shown. All conditions and interpretations are the same as for Ix1 ligation. It should be noted that the lanes ran aberrantly in this gel because a long bubble formed beneath the gel during the run. This did not affect the ability to quantitated the populations.

allowed for excessive degradation as seen in Figures 4.3, 4.4, and 4.5 for GCN4 bZip samples at times of 60 minutes and beyond. Additionally, if the putative nuclease had exonuclease activity, any digestion of the XhoI site would decrease or end the ability

of the two strands to be ligated. This degradation likely accounts for the diminished reaction rate observed in the GCN4 bZip samples as well as the overall decrease in dimer formed at the 1080 minute time point observed in Ix2 and Ix3 ligation.

Subsequent extended incubation tests (not shown) with DNA and the GCN4 bZip peptide sample confirmed that the DNA nuclease activity is an element of the protein sample and does not reflect contamination unique to this experiment.

#### *4.4 Discussion of Results*

The benefit for using a low concentration of DNA [2 nM] is that with low DNA concentration, dimerization by diffusion of DNA-only is very slow. The rate of ligation in sandwich complex is accelerated is and only if the effective concentration of the two XhoI ends near each other is greater than the bulk DNA concentration. The protein concentration was chosen empirically to maximize the dimerization product observed at the given DNA concentration of 2 nM. This allowed for better contrast among the rates of LZD73 and LZD87 bound samples and non-sandwiched samples. The Inv-2 site DNA was chosen for this experiment because the EMSA results indicated that both ends would bind to this sequence at the given concentrations. The concentration ration (protein:DNA) was increased above that of the EMSA experiments in chapter 3 based on trial and error of previous dimerization kinetics attempts (not shown). The Ix DNA alone was preferable to using a mixture of CREB and Inv-2 site DNA because it allowed for a homogeneous DNA sample. A similar set of fragments was made with the CREB site, but despite being identical in length to the Ix counterparts, they had slightly different mobilities in PAGE analysis, probably due to an initially unnoticed poly A tract in the PCR template. The use of them



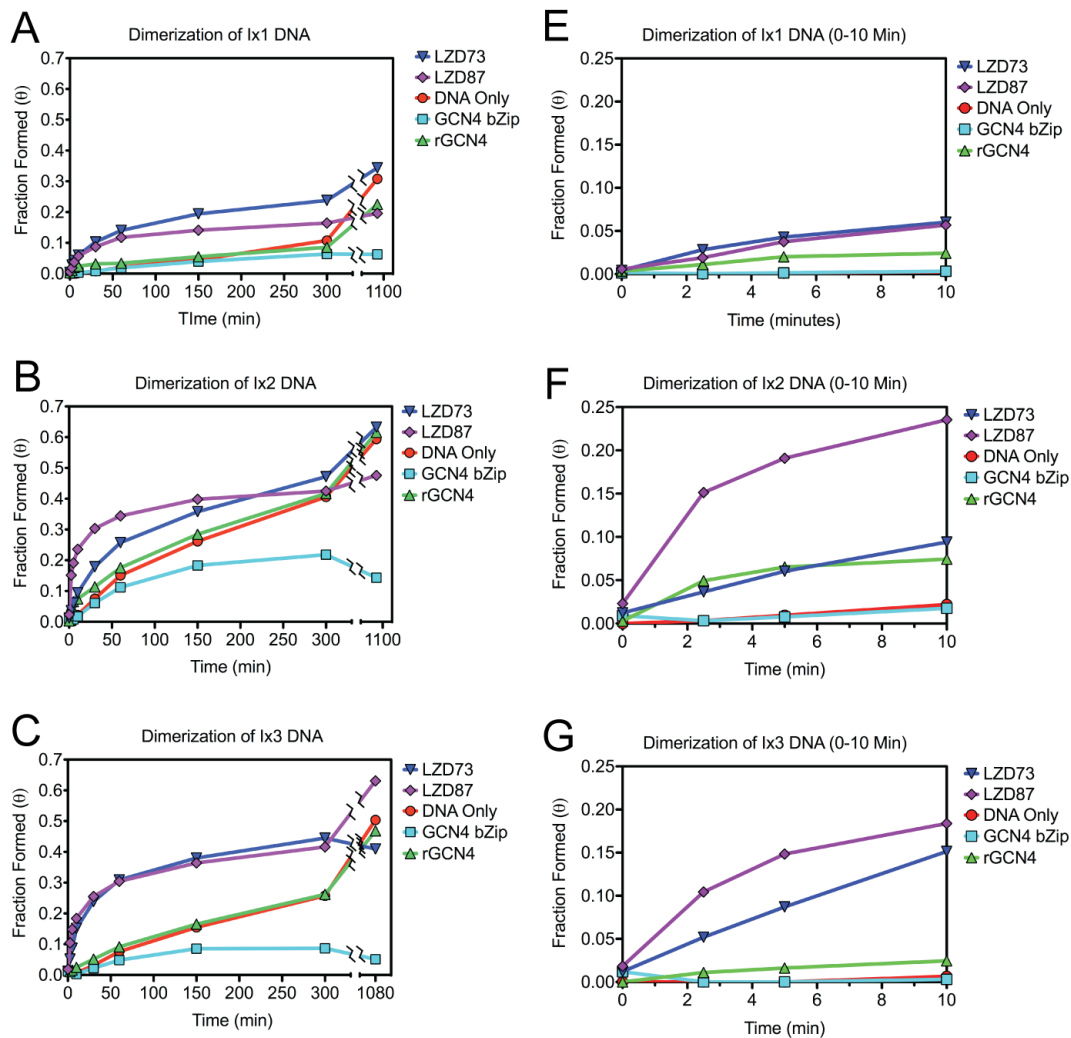


Figure 4.6 Sandwich complex-mediated dimerization kinetics. Graphical analysis of the fraction of dimer formed versus time for three lengths of DNA. The DNA samples Ix1, Ix2, and Ix3 were graphed showing all data points (A, B, and C, respectively). A second set of graphs, showing the data points for only the first ten minutes, depicts the initial enhancement of dimerization for Ix1, Ix2, and Ix3 samples (E, F, and G, respectively) with LZD 73 and LZD 87.

together resulted in different mobilities for each of the expected dimers of Ix-Dx, Ix-Ix, and Dx-Dx. Although this heterogeneous DNA experiment did show a qualitative increase in dimer product with LZD proteins (not shown), the analytical difficulties of parsing the data from three sets of products was easily avoided with the approach taken to acquire the data here.

A shortcoming of this method is that multiple protein binding conformations are possible, because the DNA is asymmetrical in length with respect to the position of the binding sites. The palindromic nature of the binding sites allows for one of two binding orientations at both ends of the protein. As seen in Figure 4.7, the sandwich complex can exist in one of two possible conformations, in either a *cis* or a *trans* state. We proposed that for shorter length DNA in the *cis* form, ligation is possible when the binding sites orient in the same direction. In the *trans* form, however,

#### Proposed Binding Orientations With Symmetrical Binding Sites

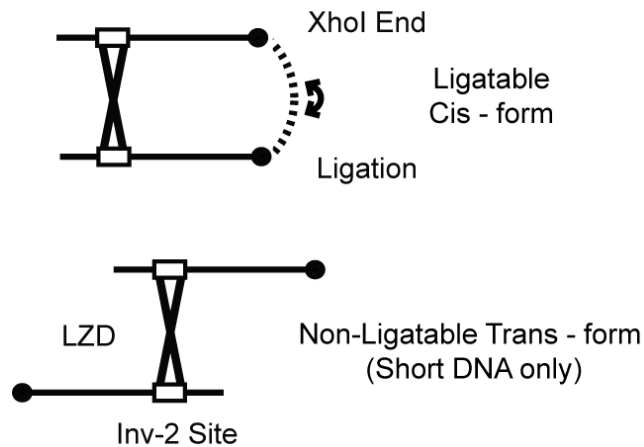


Figure 4.7 Proposed binding orientations for the sandwich complexes. If the binding orientation of a loop more closely resembles the *trans* form it may require longer lengths of DNA to observed products in the cyclization work in chapter 5.

the sites orient away from each other thus preventing ligation at short DNA lengths.

This may account for the lower overall dimerization rate observed in LZD73 and

LZD87 in the Ix1 set relative to the Ix2 and Ix3 sets. If the XhoI digest ends are

being held apart while the protein-mediated sandwich complex formation exists in the

*trans* form, ligation may be inhibited. A more likely explanation, though, is that the

quality of DNA is different between the three sets. Fortunately, this explanation does not limit the information gained from the analyzing the amount of ligated product versus time. If XhoI digestion efficiency was lower for the Ix1 DNA relative to the other two then the Ix1 DNA will have a lower effective concentration. If the DNA has more defective and therefore non-ligatable ends, there will also be an overall lower fraction formed. The support for this latter case is that the overall rates for the DNA-only control and *reverse*GCN4 are also slower with Ix1 than with Ix2 or Ix3. GCN4 bZip is of little value for comparison because of the DNA nuclease contamination.

The enhanced ligation seen for sample with either LZD73 or LZD83 is very encouraging. It can be concluded with confidence that the close proximity induced by sandwich formation leads to a drastic increase in dimerization versus DNA diffusion controlled. Again, because of the multiplicity of possible sandwich complexes, the result does not provide much quantitative information regarding the geometry and stability of the loop, only that a loop transiently existed at the moment of ligation and is thereby stable enough to exist a significant fraction of the time. This clearly warrants further evaluation of looping with LZD73 and LZD87. In the next chapter, DNA cyclization will be used to investigate looping. This sensitive technique can provide information regarding the stability of DNA loops and can detect any effect the LZD proteins may have on the topology of bound DNA.

## Chapter 5: Topoisomer Product Distribution in Protein-Bound DNA Cyclization

## 5.1 *Overview*

The evidence of transient loop formation in the dimerization experiments was a furtherance of our goal of demonstrating a stable looped complex with our engineered proteins. However, to fully characterize the protein induced looping and gain pertinent data regarding loop shape and stability, a more precise method is needed. DNA cyclization is well suited for this because it allows for differentiation between subtle topological changes. Two series of experiments were designed to first identify a minimal DNA length for looping with our proteins and then demonstrate periodicity in the results that is dependent on the helical repeat. This chapter is summarized by an upcoming manuscript submission (Gowetski, D., Kodis, E., Kahn, J. Manipulation of DNA Topology Using an Artificial DNA Looping Protein. 2012)

## 5.2 *Principles Behind DNA Cyclization in Topology Studies*

DNA ring closure (ligase-mediated cyclization) experiments have frequently been used to study the thermodynamics behind DNA flexibility in solution (Geggier & Vologodskii, 2010; Kahn, Yun, & Crothers, 1994; Rybenkov, Vologodskii, & Cozzarelli, 1997). Any particular DNA fragment, of sufficient length, with two complementary overhangs will, when exposed to T4 DNA ligase in the presence of ATP, react in one of two ways. The first is bimolecular, where a dimer product is formed, and the second is unimolecular, where a cyclized product is formed. The ratio of rates of formation of cyclized monomer circle over the formation of dimeric products is referred to as the *j*-factor and the mathematics describing this concept were actually worked out a few years before the structure of DNA was proposed (Jacobson, Beckmann, & Stockmayer, 1950). The technique was later applied to DNA by Shore and Baldwin (Shore, Langowski, & Baldwin, 1981). Recent work has used the cyclization technique to support the theory that DNA is far more flexible at short lengths than previously accepted (Cloutier & Widom, 2004; 2005). In these experiments, DNA was shown to cyclize with far higher frequency than would be expected at the given length (<110 bp), thus challenging the consensus on *j*-factors at these short lengths, although this claim has been met with skepticism (Du et al., 2005). But the approach has application beyond the determination of bending and twisting of naked DNA. As illustrated in Figure 5.1, when a linear fragment of DNA closes into a ring, any topological change, such as a change in twist or writhe, will be locked into place by the covalent closure of the ring with T4 DNA ligase. This

entrapment of form allows for a quantitative assessment of the energy differences of supercoiling (sc) behind the topological distributions ( $\Delta Lk$ ) between populations of DNA (Horowitz & Wang, 1984). It also traps a record of any structural changes induced by ligands bound to the DNA that have topological effects. This *in vitro* approach offers the potential to indentify and characterize stable DNA loops that may be formed with our proteins.

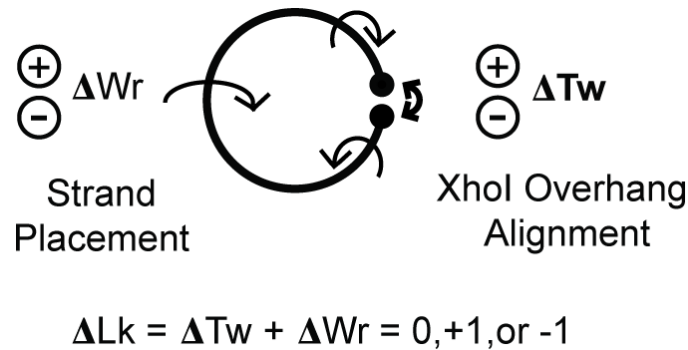


Figure 5.1 Illustration depicting the conformational changes that may lead to different topoisomer products. A change in the writhe ( $\Delta Wr$ ) can be captured if the cyclized product has two double helix strands that approach each other prior to ligation (Strand Placement). The DNA can also produce topological variants depending on whether the strand ends may under or over-twist in order to align the overhangs prior to ligation, resulting in a  $\Delta Tw$ .

When a protein binds a DNA fragment and the complex forms a loop, the protein is forcing a conformational change on the nucleic acid polymer. This change can be in the form of twist or writhe, or a combination of both depending on the relative energetics of different deformations. When the looped DNA is cyclized by DNA ligase, the conformational change in twist and/or writhe will result in a change in the linking number relative to closure without protein-induced looping (relaxed DNA). Figure 5.2 illustrates how the two physical elements can be perturbed during looping.

Recall from Section 1.2 that circular DNA of equal length but with different linking numbers can be easily resolved by electrophoresis in the presence of an intercalating agent such as chloroquine. If cyclization demonstrates an absolute change in linking number and this change is dependent on the periodicity of the helical repeat between binding sites, we can claim with confidence that the change in  $Lk$  is the result of a stable loop. If a loop can form but is unstable and quickly falls apart then the probability that ligase would cyclize a looped geometry rather than a relaxed unlooped geometry is low and the product distribution would appear unchanged from the DNA-only control. If, however, the loop is stable, then

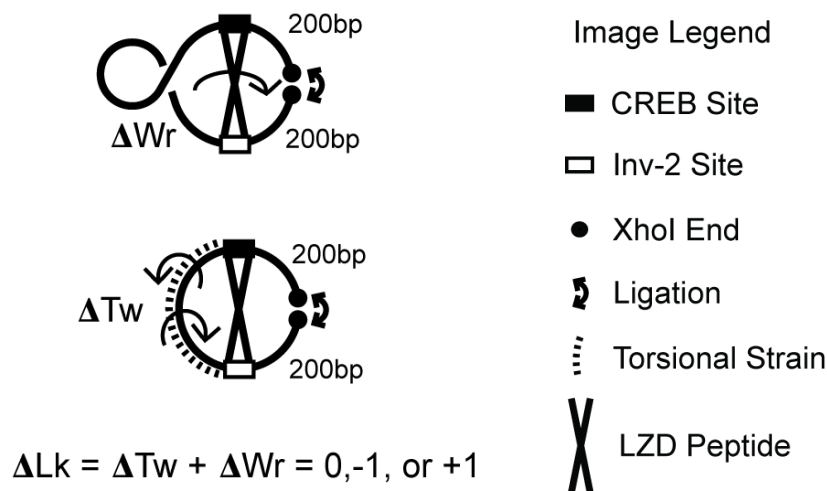


Figure 5.2 Topological changes, in the form of  $\Delta Writhe$  and  $\Delta Twist$  induced by protein looping, can manifest in different topoisomer populations. The changes in  $Wr$  and  $Tw$  result from direct geometric changes introduced by the protein and also from the bending and twisting of the unbound DNA required for loop closure. The figure legend (right) applies to all ligation-based cartoons.

it will be the most frequent form and should be readily captured by the enzymatic reaction. Additionally, a positive change in linking number would be a stronger indication of looping than a negative change. This assertion arises from previous work demonstrating that GCN4 binding induces a slight untwisting ( $53^\circ \pm 3^\circ$ ) of the DNA at the binding site that could increase the likelihood of a  $\Delta Lk = -1$  topoisomer



(Hockings, Kahn, & Crothers, 1998). It is also quite possible that both a +1 and -1 topoisomer will result because a stable loop may form by either undertwisting (-1 topoisomer) or by overtwisting (+1 topoisomer) the DNA between the protein's two binding sites. Though a topological change should arise from looping, it is not proof of a DNA loop.

The telltale sign of looping, as demonstrated by the *in vivo* work of Tom Record and Benno Müller-Hill (LacI), as well as Robert Schleif (AraC), is to demonstrate that a given result is dependent on the periodicity of the helical repeat (Bellomy et al., 1988; Lobell & Schleif, 1991; Müller et al., 1996).

In order to fully characterize the looping capacity of the LZD proteins, we designed two sets of cyclization experiments. The first investigated the loop stability, as measured by the appearance of topological changes, over a range of DNA fragments where the binding sites were spaced between 153-448 bp. The second sought to demonstrate a periodic dependency in the resulting topoisomer changes by cyclizing a series of DNA fragments of equal length that contained incremental spacing variations between the CREB and Inv-2 sites over two helical repeats (435-458 bp). Combined, these two methods were used to provide a quantifiable approach to characterizing DNA looping geometry induced by the LZD proteins.

### 5.3 *Materials and Methods*

#### 5.3.1 Design of Variable-Length DNA Constructs For Cyclization

A series of DNA fragments of variable lengths was constructed to determine whether stable loops could be formed using the LZD proteins. Figure 5.3 shows the

construction of the plasmids used in this experiment and the primers used. Six plasmids were constructed to contain both the CREB binding site (5'-ATGACGTCAT-3') and the Inv-2 binding site (5'-GTCATATGAC-3') separated over a range of 153-448 bp. PCR was used to synthesize cyclization fragments that could then be digested with XhoI at both ends to produce two complementary overhangs that would permit cyclization. Two sets of primers were used to allow two possible lengths from the binding sites to the XhoI ends. This translates to the effective size of what we have referred to as the “outer loop”, the size of the DNA fragment between the two binding sites that is covalently closed by T4 DNA ligase. The distance between the CREB and Inv-2 sites, the DNA involved in looping, we refer to as the “inner loop”. Primers were designed to produce external loops of 212 bp (primers 100 and 101) or 414 bp (primers 200 and 201). Complete sequences for the PCR generated fragments and primers used can be found in Appendix 1. Table 1 lists the DNA fragments used and provides a breakdown of the overall lengths and binding site spacing separation.

DNA Construct	Length (bp)	Helical Turns (Lk°)	Binding Site Separation (bp)
Vx(153)200	567	54.0	153
Vx(202)200	616	58.7	202
Vx(254)200	668	63.3	254
Vx(310)200	724	69.0	310
Vx(376)200	790	75.2	376
Vx(448)200	660	62.8	448
Vx(448)100	862	82.1	448

Table 1 DNA fragments constructed by cloning, PCR and XhoI digestion. The total length, number of helical turns (N/10.5 bp), and the binding site separations are given.

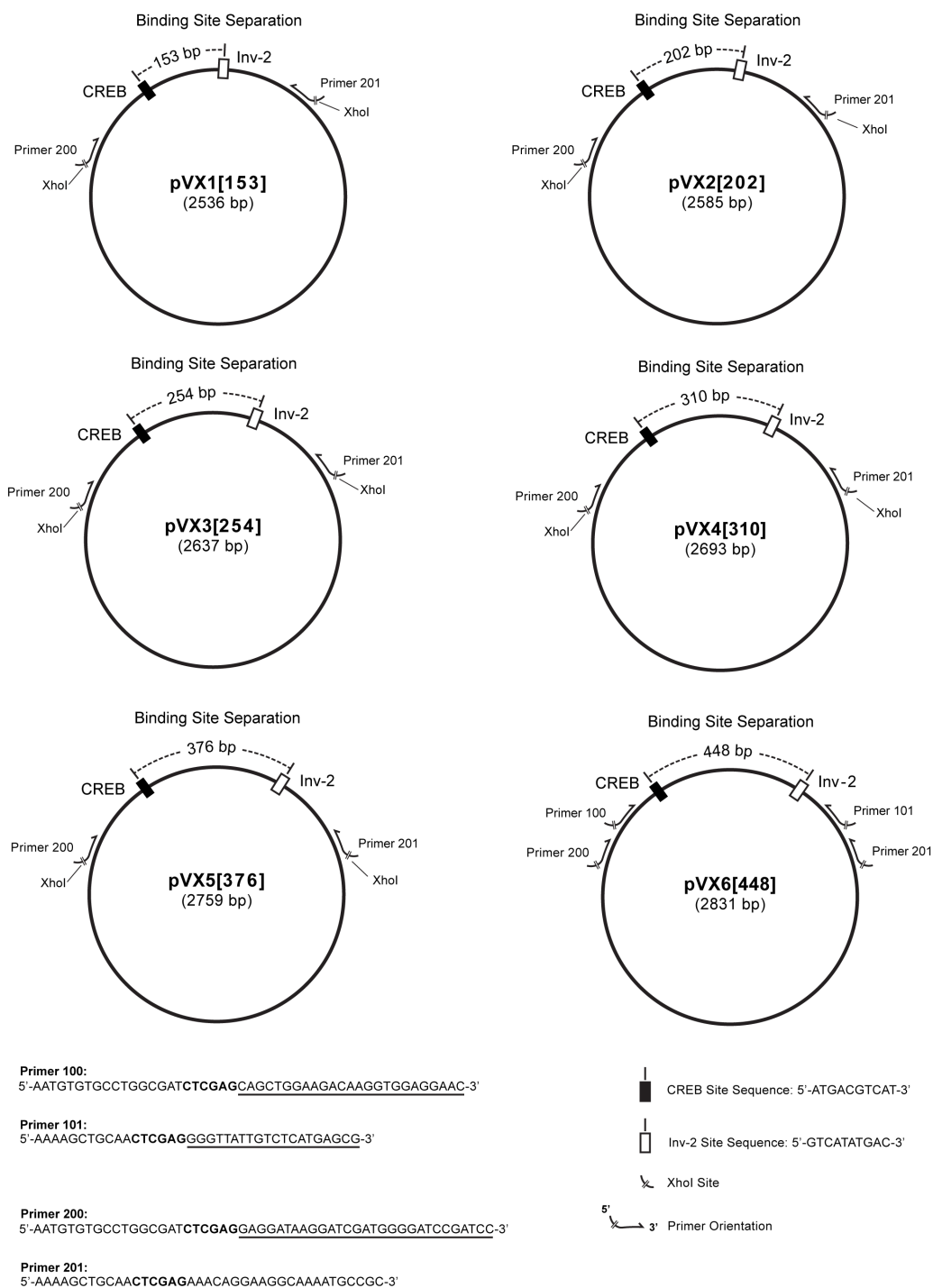


Figure 5.3 Plasmids used for producing the Vx(153-448) DNA fragments for cyclization. CREB sites are identified as solid black boxes, while Inv-2 sites are white boxes. PCR followed by XhoI digestion of the products will produce DNA fragments with variable inner loops of 153-448 bp and constant outer loops of 212bp (primers 100 and 101) or 414 bp (primers 200 and 201).

### 5.3.2 Design of Vx 435-458 Binding Site Separation Fragments

To demonstrate the periodic dependence of the topoisomer populations on the helical phasing of the binding sites, a series of 10 plasmids were constructed that contained both the CREB binding site (5'-ATGACGTCAT-3') and the Inv-2 binding site (5'-GTCATATGAC-3') separated incrementally over a range of 435-4548 bp. Primers 200 and 201 from the previous PCR protocol were used to incorporate the two XhoI sites at the ends of the PCR products. Figure 5.4 illustrates the plasmid

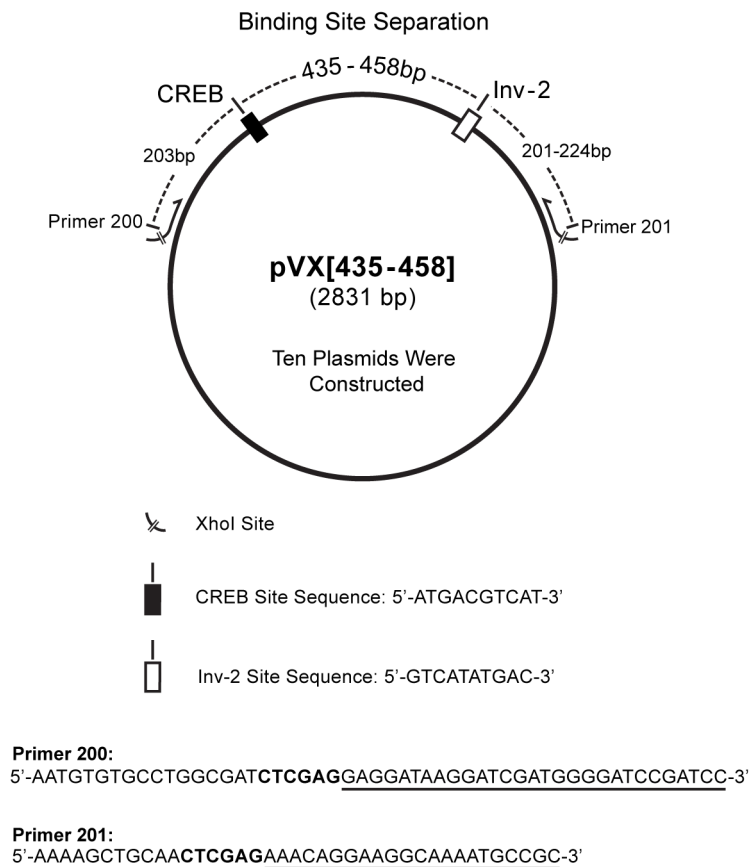
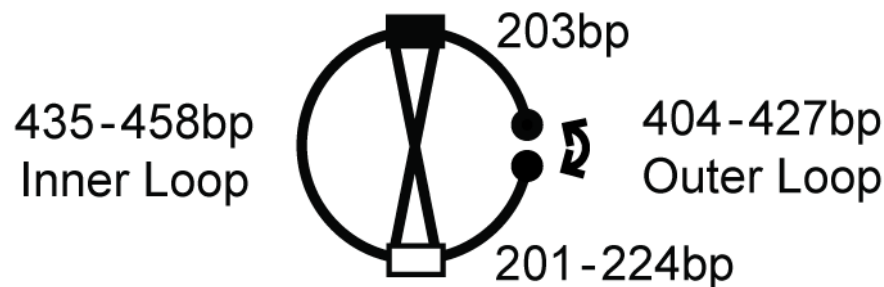


Figure 5.4 Schematic of the ten plasmids used to construct the Vx(435-458) DNA fragments by PCR. CREB sites are identified as solid black boxes, while Inv-2 sites are white boxes. PCR with identical primers followed by XhoI digestion of the products will produce DNA fragments with a constant length of 862 bp with variable inner loops of 435-458 bp and variable outer loops of 427-404 bp. Underlined regions of the primers represent the complementary sequence to the template while the XhoI site is in **BOLD**.

design and placement of the primers with respect to the binding sites. In order to maintain a constant length for the ten fragments (so that the DNA-only controls cyclize identically), we varied the number of base pairs between the Inv-2 binding site and XhoI end. The “internal loop” (between the binding sites) varied



$$\Sigma \text{ Inner} + \text{Outer} = 862\text{bp Cyclized Product}$$

Figure 5.5 Cartoon illustration depicting the Vx(435-458) phased binding site experiment. The variable length inner loop and outer loop are illustrated. Total lengths for these ten fragments was held constant at 862 bp.

between 435-458 bp. To maintain a constant length of 862 bp for all ten fragments, the Inv-2 to XhoI length was varied from 224-201 bp, respectively, while the length from the CREB site to its corresponding XhoI end was held constant at 204 bp.

DNA Construct	Length (bp)	Binding Site Separation (bp)	Internal Helical Turns (Lk°)	External Helical Turns (Lk°)
Vx(435)200	862	435	41.4	40.7
Vx(438)200	862	438	41.7	40.4
Vx(440)200	862	440	41.9	40.2
Vx(443)200	862	443	42.2	39.9
Vx(445)200	862	445	42.4	39.7
Vx(448)200	862	448	42.7	39.4
Vx(450)200	862	450	42.9	39.2
Vx(453)200	862	453	43.1	39.0
Vx(455)200	862	455	43.3	38.8
Vx(458)200	862	458	43.6	38.5

Table 2 DNA fragments used in Vx(435-458) cyclization. The total length is held constant at 862 bp while the binding site is spaced between 435 and 448 bp. The helical repeats of both the “inner loop” (between CREB and Inv-2 sites)) and “outer loop” (formed upon ligation) are given.

The constant length of 862 bp (82.1 helical turns) and the sequence near-identity between fragments differing only in the placement of the Inv-2 site should ensure that these DNA fragments behave identically in the DNA control reactions. Table 2 lists the 10 fragments used in this experiments. The full sequences of the PCR products can be found in Appendix 1.

### 5.3.3 Assembly, Radiolabeling, and Purification of DNA Constructs

PCR generation of  $^{32}\text{P}$  labeled DNA fragments was done using PCR mixture protocols as described in the PCR reactions described in the Materials and Methods section of Chapter 4, except scaled for more samples. Each PCR reaction was performed at 50  $\mu\text{L}$ , though the master mixes varied in volume depending on the number of total reactions performed. Additionally, the PCR cycle parameters were modified to achieve better yield as follows:

Initial denaturation: 99 °C for 3 min

Cycle (33 repeats)

1. 95 °C for 1 min
2. 65 °C for 20 sec
3. 73 °C for 25 sec

The PCR products were subsequently purified by Qiaquick PCR cleanup kit (Qiagen), eluted in  $\text{H}_2\text{O}$  (50  $\mu\text{L}$ ), and digested overnight with 20 units of XhoI enzyme (NEB) in NEB Buffer 2 (57  $\mu\text{L}$  total volume). The samples were then gel purified using 7 % acrylamide (75:1), extracted and further processed by Qiaquick, prior to determining the concentration by scintillation counting as described in the

DNA purification section of Chapter 4 Materials and Methods. The final DNA products were in H<sub>2</sub>O after elution from the second Qiaquick PCR cleanup column and were stored at -20 °C.

#### 5.3.4 Ligase-Mediated Cyclization of Protein-Mediated DNA Loops

The buffer conditions used in the experiments were identical to those described in the Materials and Methods section for the EMSA experiments in Chapter 3 and the dimerization assay in chapter 4. All samples were diluted in and all reactions took place in B/L buffer. The stop buffer used consisted of Proteinase K (NEB) diluted to 2 mg/mL in B/L buffer that had been brought to 50 mM EDTA to chelate the Mg<sup>2+</sup> and quench ligase (EDTA was 10 mM upon addition to the reaction).

DNA samples were diluted to 1 nM and protein samples were diluted to 30 nM in B/L Buffer. T4 DNA ligase was diluted to 10 U/μL immediately prior to use. The proteins used in the variable length (153-448 bp) experiment were the *reverseGCN4* peptide, LZD73, and LZD87. GCN4 bZip was not used because of the putative DNA nuclease contaminant issues observed during the dimerization assays. The use of DNA and *reverseGCN4* single binding protein as controls was deemed sufficient. The protein and DNA samples were mixed in equal volumes (5 μL each) and allowed to equilibrate for 10 minutes at room temperature. The DNA-only samples (5 μL) were mixed with 5 μL B/L buffer to maintain consistent concentration. T4 DNA ligase diluted in B/L buffer (10 U/μL) was then added in equal volume (5 μL) to the samples and DNA and mixed by gentle pipetting. The final concentrations during ligation were 0.33 nM DNA, 10 nM LZD or control

protein, and T4 DNA ligase at 3.3 U/ $\mu$ L. The reaction was allowed to proceed at room temperature for 60 minutes. Following ligation some samples were treated with BAL 31 DNA nuclease to remove linear DNA from the sample to remove any linear multimers background bands that could interfere with analysis of topoisomer products. For these samples, 15  $\mu$ L of 2X Bal-31 reaction buffer was added (2X BAL 31 buffer is 40 mM Tris HCl, 1.2 M NaCl, 24 mM CaCl<sub>2</sub>, 24 mM MgCl<sub>2</sub>, 2 mM EDTA pH 8.0) and then 0.25 units of BAL 31 was added to each sample followed by mixing by pipetting. The reaction was allowed to proceed at 30 °C for 30 minutes. Following this digestion step, 4  $\mu$ L of 2 mg/mL Proteinase K in B/L with 50 mM EDTA was added and samples were moved to 37°C for 15 minutes. Samples that were not BAL-31 digested had 4  $\mu$ L of the Proteinase K mix added immediately after the 60 minute ligation and the samples were also incubated at 37 °C for 15 minutes to digest the T4 DNA ligase and target proteins. These samples then had 15  $\mu$ L of the 2X BAL-31 buffer added (but no enzyme) to increase the ionic strength of the solution so it would be suitable for EtOH precipitation, and the make sure that salt concentrations were consistent in later steps.

#### 5.3.5 EtOH Precipitation of Reacted Samples and Gel Analysis

All samples were EtOH precipitated prior to gel analysis. To each sample (either 30 or 35  $\mu$ L at this point), 105  $\mu$ L of 100 % EtOH was added and the samples gently mixed by pipetting. The samples were moved to the -80 °C freezer for 15 minutes. They were then centrifuged for 15 minutes in the micro centrifuge at 16,000 x g at 4 °C. The supernatant was removed and the pellets were air dried for 3 minutes. The samples were then suspended in 15  $\mu$ L 1X DNA loading dye (0.05 %



bromophenol blue, 0.05 % xylene cyanol, 3 % Ficoll 400, and 10 % glycerol (without the high glycerol these samples tended to “float” upon gel loading), and chloroquine (7.5 µg/mL) intercalator to resolve the topoisomers. The samples were moved to a 50 °C bath for 5 minutes prior to loading to improve resuspension after EtOH precipitation.

Gels consisted of 6 % acrylamide (75:1) in 50 mM Tris HCl, 50 mM boric acid, 1 mM EDTA, 7.5 mM NaCl, and 7.5 µg/mL chloroquine. The identical buffer formulation was used as the running buffer. Gels were run at 4 V/cm for 18 hours (Vx153-448 samples) or 42 hours (Vx435-458 samples). The gels were dried and then exposed to a storage phosphor screen for at least 24 hr. The images were captured using the Storm Phosphorimager 860. The bands were quantitated using the volume integration function on ImageQuant (Molecular Dynamics). For the Vx435-458 samples, the fraction formed for each topoisomer population was calculated by dividing each Lk population by the sum of all Lk populations.

The values were plotted as a function of binding site separation to illustrate the correlation between topoisomer populations and the helical repeat of DNA using Prism 5.

## 5.4 *Results*

### 5.4.1 Formation of Topological Variants for Variable Length DNA

#### Constructs Vx(153-448)

A unique topological shift can be viewed as evidence for the formation of a loop, especially if the new topoisomer is of the +1 variant, as explained in the

Introduction. However, there are many alternate reactions that may occur leading to other final products. The first consideration is that even in a system that involves DNA looping, it would be unrealistic to assume that all DNA molecules are bound in a looped state. The concentration of protein may not be adequate to ensure complete binding saturation, and any loop will be in equilibrium between a looped and unlooped state. If ligase were to react with the DNA that either was not bound to protein or only singly bound, the topoisomer product should loop identically to the DNA-only control. The second element to consider would be the formation of a sandwich complex ( $P \cdot D_2$ ) that could react in one of two ways. As illustrated in Figure 5.6, this event would lead to a bimolecular product or cyclization that is independent of the bound protein. For DNA fragments that cannot loop, this would be the predicted result. In Figure 5.7, the bimolecular products are enhanced to a degree in nearly every LZD73 or LZD87 sample. However, the effect is greatest in the Vx(153)200 sample and the Vx(448)100 sample. In these samples, it is also observed that the circular products are diminished relative to the DNA-only control

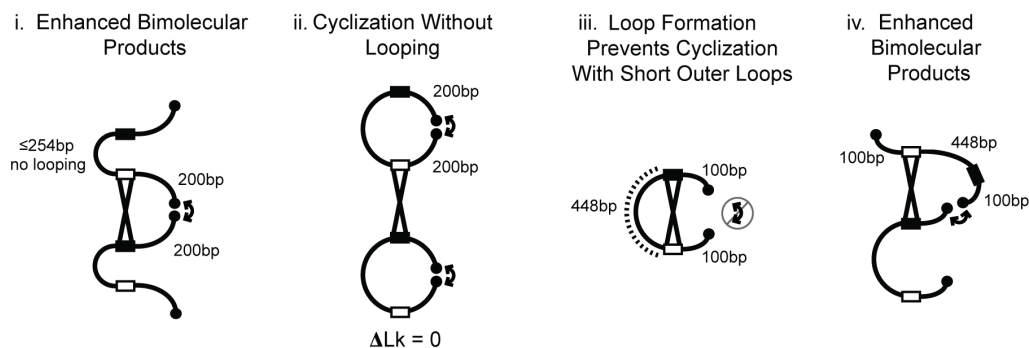


Figure 5.6 Reaction outcomes that do not give novel topological products: i. Sandwich complex formation leads to enhanced dimerization ii. Protein independent cyclization iii. Cyclization is inhibited by looping. iv. Enhanced bimolecular reaction through alternate geometry sandwich complex formation. Enhanced bimolecular reaction can also occur through kinetic partitioning if the rate of cyclization decreases in a looped complex

because they are being diverted to bimolecular products. Figure 5.6 (right) illustrates the possible products associated with this control fragment with a shortened external loop.

What emerges from the results in Figure 5.7 is that the DNA fragments with binding site separations  $\leq 254$  bp did not result in the formation of new topoisomer products in the presence of LZD proteins (lanes d, e, and f in columns 1, 2, and 3) suggesting that looping does not occur at these lengths. The samples with  $\geq 310$  bp

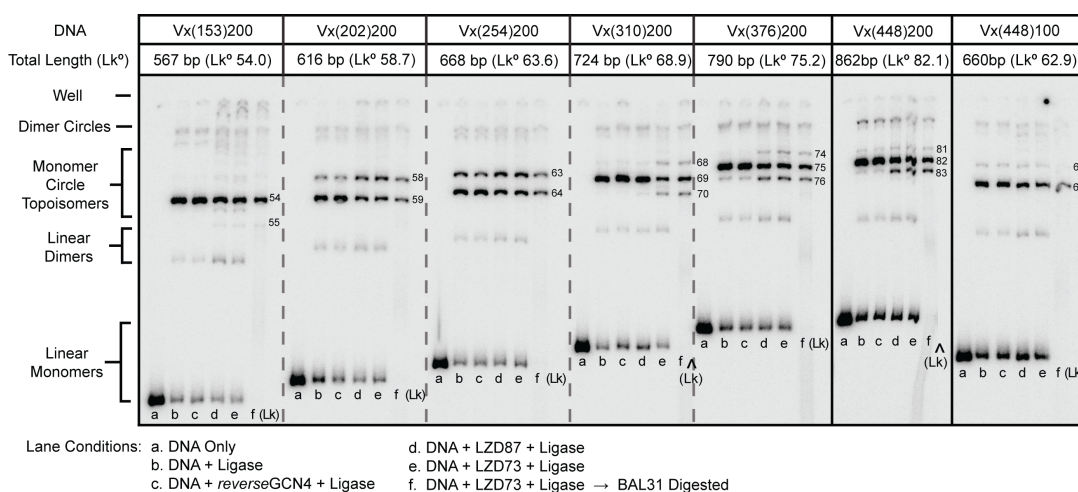


Figure 5.7 Analysis of Vx153-448 cyclization constructs run on a 6 % acrylamide gel with 7.5 µg/mL chloroquine. The repeating lane conditions (a-f) are as follows: a. DNA-only, b. DNA + ligase control, c. DNA + *reverseGCN4* + ligase control, d. DNA + LZD87 + ligase, e. DNA + LZD73 + ligase, f. DNA + LZD73 + ligase then BAL-31 digested. The presence of +1 and -1 topoisomers in the LZD added samples (e, f, g) for DNA  $\geq 310$  bp is indicative of looped DNA.

binding-site separation, however, all demonstrated new topoisomer products in the form of both +1 and -1 topoisomer products (lanes d, e, and f columns 4, 5, and 6). Additionally, the shortened external loop control (condition iii in Figure 5.6, column 7 of Figure 5.7) did not result in new topoisomer products. We propose that a loop does form between 448 bp as with the Vx(448)100 fragments but the XhoI ends are

oriented away from each other and cannot ligate (Figure 5.6 iii.), supporting the theory of the *cis* and *trans* orientation is correct (Section 4.3).

The emergence of +1 topoisomers during the cyclization of fragments where the binding sites are spaced  $\geq 310$  bp is a very strong indication of protein-mediated looping. If only -1 topoisomers emerged, that could be explained by local untwisting at the point of single binding with protein, while the +1 population cannot be a result of such a binding event. These results represent a closer step towards a definitive demonstration of looping, but they are not conclusive in their own right. To provide an unequivocal answer regarding loop formation with our proteins, the 435-458 bp binding-site phasing experiment will be essential.

#### 5.4.2 Optimized Protein Concentration for Looping

Protein concentration plays a pivotal role in determining the outcomes in these cyclization assays. The optimal concentration of protein is one where all, or nearly all of the DNA is looped and very little of it is free in a sandwich complex with protein, or doubly-bound by protein. If the protein concentration was below the optimal value then there would exist an unbound portion of the DNA population that would be cyclized with the same distribution as the DNA-only controls. This would effectively lead to a high degree of background noise in the sample distribution. Although this would be uniform throughout the samples (provided the protein concentration was held constant) it is still an undesirable and avoidable scenario. In contrast, a protein concentration that was well above the optimal level would lead to non-specific binding and double-bound complex formation. At very high protein concentrations, network complexes would emerge, essentially polymerizing the

protein:DNA complexes. The addition of T4 DNA ligase to such a complex would lead to not only enhanced dimerization, but also to multimers of DNA including trimers, tetramers, and beyond.

To evaluate how protein concentration affects the distribution of cyclized versus multimerized products, Vx(448) DNA was cyclized in a series of reactions where LZD73 concentration was incrementally increased from 4-66 nM while the DNA was held constant at 0.33 nM (DNA concentration was kept low to minimize sandwich complex formation and reduce bimolecular ligation).

The gel analysis for protein concentration optimization (Figure 5.8) demonstrates an initial increase in the +1 topoisomer (the determinant product) at protein concentrations from 4 nM to 10 nM. This population then begins to decrease

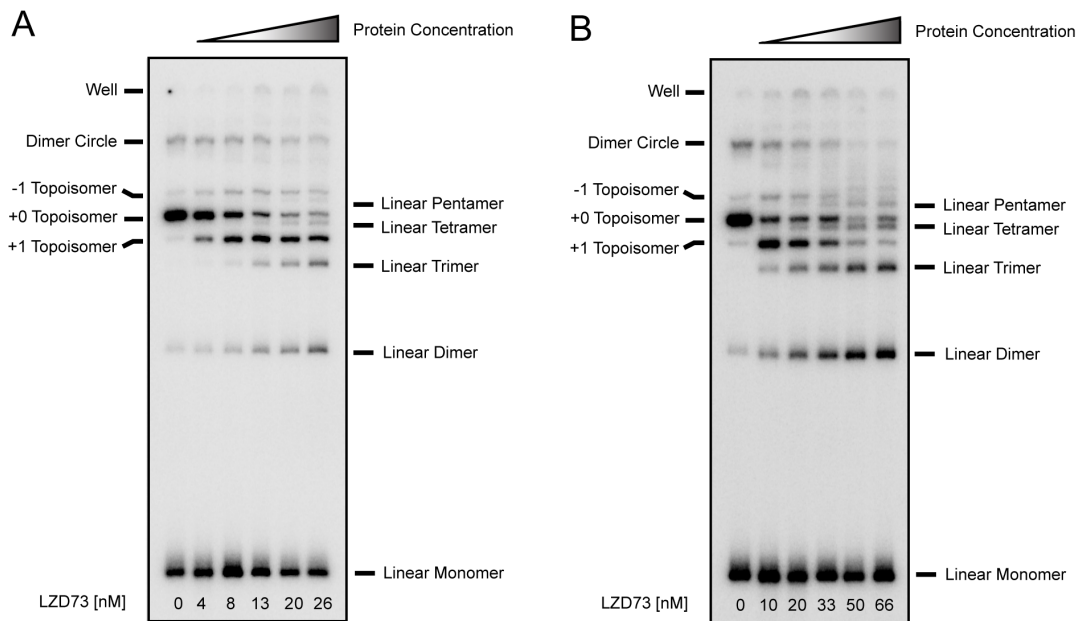


Figure 5.8 LZD73 protein gradients from 0-66 nM combined with 0.33 nM Vx6(448)200 and T4 DNA ligase. The +1 topoisomer is the determinant variable for observing looping effects. This population increases up to 10 nM and then steadily decrease with LZD concentrations above that. The linear products (dimer trimer tetramer) steadily increase with increasing protein concentration. Reactions were performed for 45 minutes at 3.3 U/uL T4 DNA ligase in B/L buffer.

as the LZD73 concentration increases further. It is also abundantly clear that as the concentration of LZD73 increases, the amount of linear dimer, linear trimer, and linear tetramer greatly increase and eventually become the dominant populations formed. Since the linear products are evidence of sandwich complexes (likely non-specific at high concentrations) and compete with or outright inhibit cyclization, their emergence is detrimental to this assay. It can be concluded that the 10 nM protein sample maximizes the +1 and -1 topoisomers, which are the desired populations that correspond to looped DNA. Therefore the reactions of the Vx(435-458) binding site phasing were performed at this concentration in good confidence that it is the optimal condition.

#### 5.4.3 Variable Length Vx(435-458) Formation of Topological Variants

Following the classic *in vivo* work with the LacI that demonstrated looping by showing a correlation between activation or repression and the helical phasing of binding sites, the Vx(435-458) experiment sought to leave no doubt regarding the ability of the LZD mutants to loop DNA. The decision to shift the binding sites incrementally over two helical repeats around the Vx(448) DNA fragment was not arbitrary. While the results from the previous experiment suggested that looping could occur at 310 bp, a more intense signal was observed with the Vx(448) sample. Of additional significance was the corresponding size of the external loop formed at the moment of ligation. It seemed a more prudent step to try to keep the internal and external loop of comparable size (435-458 bp versus 427-404 bp), because modeling of DNA energetics is thereby much simpler.

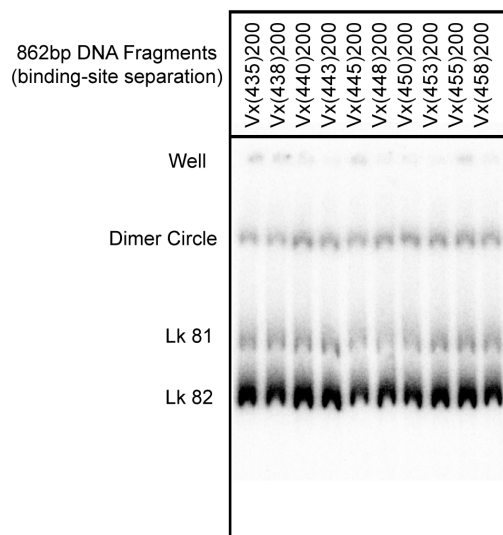


Figure 5.9 DNA-only cyclization of Vx(435-458) constructs with XhoI cohesive ends. All DNA are 862 bp with near identical sequences (see Appendix 1). 6 % acrylamide gel with added chloroquine (7.5  $\mu\text{g/mL}$ ) allows for separation of topoisomers. The observed topological products are assigned to Lk 82 (major) and Lk 81 (minor). All reactions were treated with BAL 31 before electrophoresis.

If this result is to have any merit it must be shown that the DNA-only controls do not result in any periodic changes in topoisomer distribution. Because the lengths are all held constant at 862 bp, in the absence of and sequence-dependent bends or large changes in composition, all molecules should give identical patterns of topoisomer distribution. Figure 5.9 shows the gel analysis of the 10 DNA-only controls in this experiment.

Graphical analysis of the DNA-only topoisomer distribution versus binding site separation (Figure 5.10) shows that the data can be fit to horizontal lines. For these 862 bp DNA fragments cyclized in B/L buffer it was found that the topoisomer populations form such that Lk81 accounts for 22 % of the total population, Lk82 accounts for 78 % of the total population and Lk83 accounts for <0.5 % of the total population. The average linking number ( $Lk_{ave}$ ) for these or any individual sample

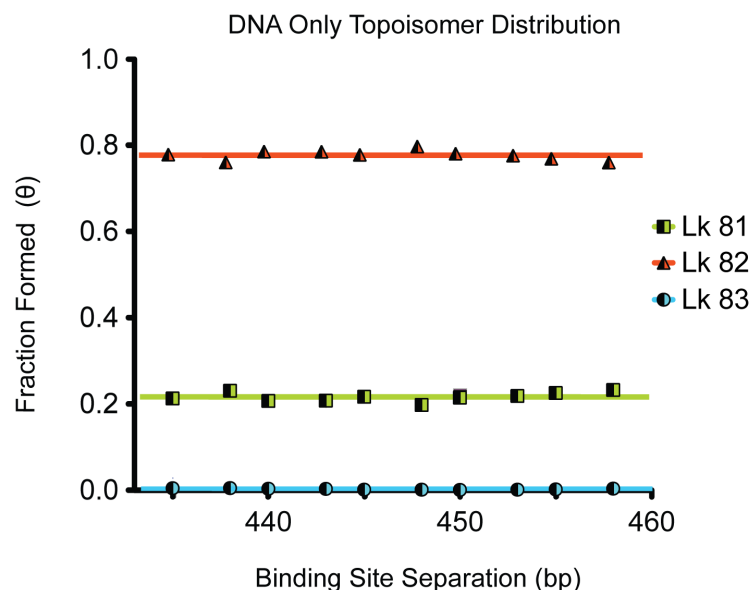


Figure 5.10 The fraction of each topoisomer formed for each of the DNA-only cyclization reactions for Vx(435-458). The data sets can be fit to horizontal lines, confirming that these DNA fragments behave identically when cyclized by T4 DNA ligase.

is determined as the sum of the products of linking number values (Lk) by the fraction of the population at that Lk. For DNA-only samples, this value is calculated to be 81.76 by the following equation:

$$Lk_{ave} = 81 * 0.218 + 82 * 0.778 + 83 * 0.004 = 81.76$$

The addition of protein to the cyclization reaction produced a dramatic effect. As seen in Figure 5.11, the topoisomer distribution expands to four topoisomers as opposed to two observed for the DNA-only controls. The expansion of observed topoisomers from two to four, however, is not the sole significance of this gel. When comparing the topoisomer distribution to the binding site separation, a pattern emerges. Two DNA-only samples (Vx(448)200 and Vx(455)200) were included in this gel (lanes 1 and 2) for comparison. This experiment was repeated (LZD73 n=5, LZD87 n=3) and the averaged topoisomer populations are plotted versus binding site



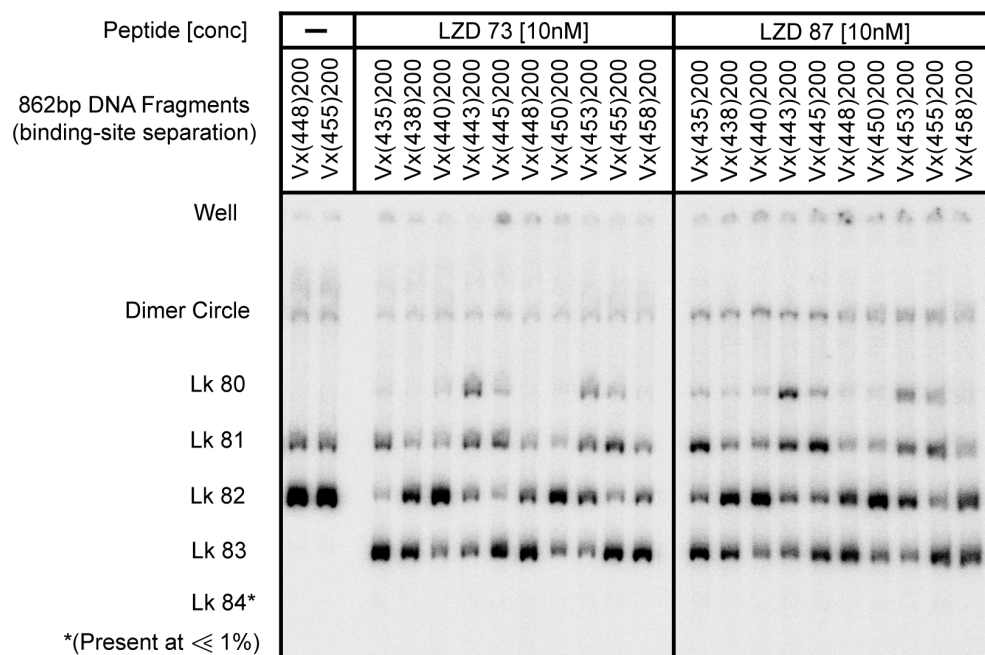


Figure 5.11 Protein-induced distribution of topoisomers. Cyclization of Vx(435-455) 862 bp DNA constructs after addition of LZD73 and LZD87 protein. 0.33 nM DNA and 10 nM protein is allowed to equilibrate for 10 minutes prior to ligation with T4 DNA ligase. Digestion of linear DNA with BAL 31 leads to only cyclized product being present in the 6 % acrylamide gel with 7.5  $\mu\text{g/mL}$  chloroquine. The ligation of looped DNA captures the equilibrium as a distribution of topoisomers.

separation as seen in Figures 5.12 (LZD73) and 5.13 (LZD87). This representation clearly illustrates the helical dependence of the topoisomer distribution. These values were fit to a sum of Gaussians using Prism, and the peaks between each LK population were found to be separated by  $10.5 \text{ bp} \pm 0.5 \text{ bp}$ . Figure 5.12 depicts the results for LZD73 and Figure 5.13 depicts the results for LZD87. The periodicity seen in the topoisomer populations versus binding site separation represents a clear demonstration of looping and validates the function of the LZD proteins as artificial DNA looping proteins.

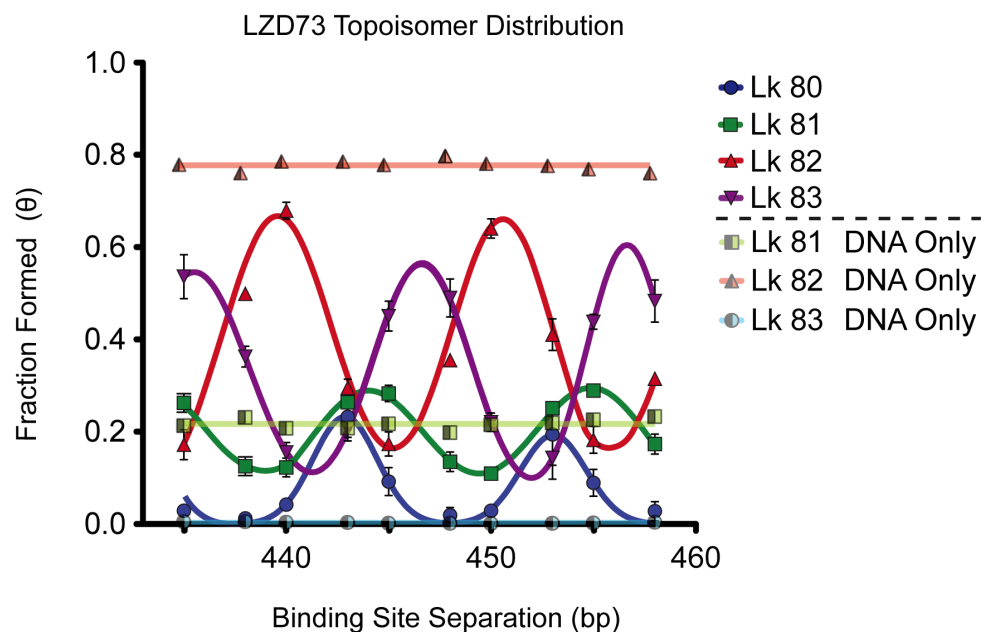


Figure 5.12 Periodicity of topoisomer distributions for LZD73-mediated loops as captured by T4 ligase-mediated cyclization of 862 bp DNA ( $n=5$ , error bars indicate the standard deviation). The DNA-only samples were repeated from Figure 5.10 for comparison. The observed periodicity is fit to a sum of Gaussians and the distance between peaks correlates to the helical repeat of DNA (10.5 bp).

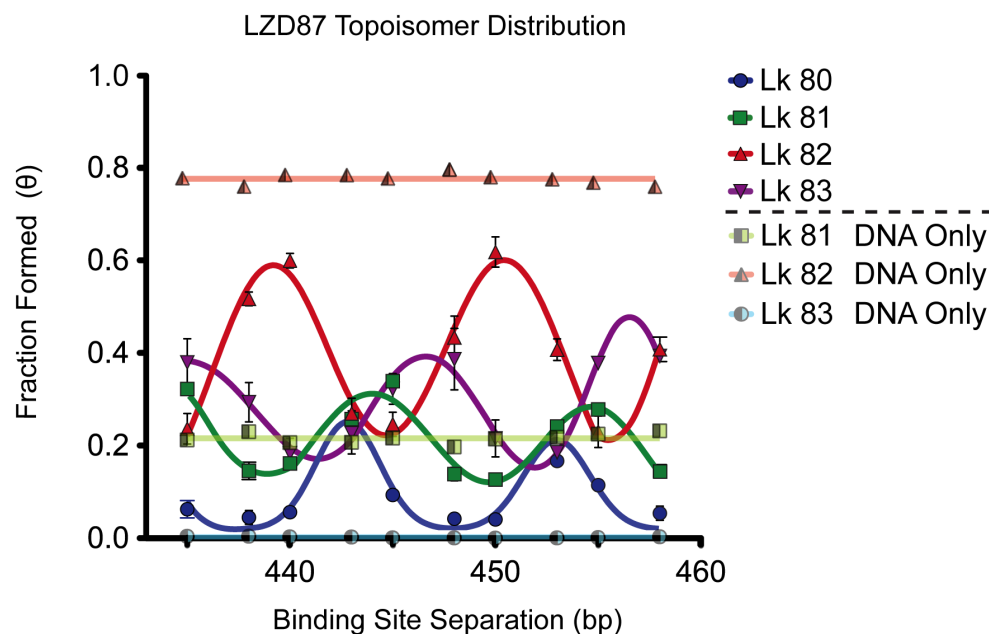


Figure 5.13 Periodicity of topoisomer distributions for LZD87-mediated loops as captured by T4 ligase-mediated cyclization of 862 bp DNA ( $n=3$ , error bars indicate the standard deviation). The DNA-only samples were repeated from Figure 5.10 for comparison.

## 5.5 *Discussion of Results*

In order to clearly demonstrate looping with our artificial proteins, we devised a series of cyclization experiments where DNA looping would alter the topological distribution of the cyclized products. The inner loop length spacing variation in the Vx(153-448) experiment sought to identify a minimal looping length range. By demonstrating both  $\Delta Lk = +1$  and a  $\Delta Lk = -1$  populations at lengths  $\geq 310$  bp, we were able to show that our protein can be used to form DNA loops at those lengths. The  $+1$  topoisomer production was a clear result of a looping protein inducing structural change on the DNA. If the  $\Delta Lk$  were simply a result of DNA untwisting upon single binding, then only  $-1$  topoisomers would form rather than both  $+1$  and  $-1$ . In order to unequivocally demonstrate looping as opposed to protein-induced overtwisting, we used an approach of phasing the binding sites over two helical repeats to show that looping produced a topological change that was dependent on helical phasing. The formation of a loop and then subsequent cyclization of this looped DNA fragment resulted in a distribution of topoisomers that can only be explained by protein induced changes in DNA writhe ( $\Delta Wr$ ) and twist ( $\Delta Tw$ ). The topology experiment also sets limits on possible loop geometries, as described below.

### 5.5.1 Contribution of $\Delta Tw$

The Vx435-458 experiment surveys the topological fate of looped DNA across the span of 24 bp. Because this represents just over two helical repeats, the observed pattern clearly repeats itself. This is reaffirming. Two of the most noteworthy topological distributions to emerge in this pattern of results are those that appear at the 440/450 repeat and those at the 435/445/455 repeat. The 440/450 set is

interesting because it so closely resembles the DNA-only sample. This could be explained in one of two ways. One is that this particular loop is inherently unstable and what is observed is simply DNA-only being cyclized by ligase. The other is that the DNA is looped and that the geometry of that loop most closely resembles that of relaxed DNA. The evidence suggests the latter. For starters, the looped DNA on each side of the protein is roughly three persistence lengths long, which means that the torsional strain of a twist of  $180^\circ$  or less in a 862 bp fragment is relatively low. The DNA-only sample exists as two topoisomers (Lk 81 and Lk 82) when fully relaxed, which suggests there is enough torsional freedom to account for up to a full twist in 862 bp DNA. Additionally, the difference between these samples and their nearest neighbors in the series is only two and three base pairs. If a loop is stable at 438 bp and at 443 bp, is hard to argue that it cannot accommodate the change in twist between the two. Furthermore, a casual glance at the pattern shows that the topoisomer distribution is shifting from heavily +1 to -1 and -2 from 438 to 443 bp. That this distribution aligns perfectly with the Gaussian fit substantiates the claim that this is not an unlooped outlier among the data pool. The final piece of evidence for this set is that there is a near integer number of helical repeats between the binding sites (41.9 and 42.9 for 440 and 450, respectively). This suggests that the DNA does not have to twist to align the binding sites as the two approach each other during looping. That the variance of topoisomers emanate away from the two data sets is strong evidence that, in these looped populations, the DNA must either over or under twist in order to align the binding sites.

The second highly interesting data sets, the 435/445/455 repeats, also support the argument that  $\Delta$ Twist plays the dominant role the topoisomer distribution. The support for a  $\Delta$ Twist theory of distributions comes from the bifurcation of the topoisomers into Lk 83 and Lk 81 at these positions. Relative to the 440/450 samples, whose internal helical repeat tally is nearly integral, the 435/445/455 set is offset by 5 bp, which translates to a helical repeat tally that is nearly half-integral. Because the binding sites now face opposite sides of the double helix, as the loop forms, the DNA must either over-twist or under-twist in order to align the DNA binding sites with the protein. That the DNA must make assume one of those two conformations is further evidence for this distribution being a result of twist variations.

The 435/445/455 set also uniquely provides evidence that the protein concentrations used in this experiment were optimal. Because Lk82 is the major component of the DNA-only, relaxed cyclization, any DNA from the protein-containing samples would be likely form Lk82 if not looped around protein. That the Lk82 population is nearly non-existent in LZD73 and very minor in LZD87, suggests that nearly all DNA was involved in protein binding. It is clear that all of the protein-bound loops have substantially DNA deformation higher free energy than relaxed because they all deviate away from the most relaxed topoisomer of 82. The DNA deformation is paid for by the favorable free energy of protein binding. However, this is not to say that all DNA is involved in looping. These samples were BAL 31 treated which digested any linear DNA so any dimer or trimer would not have been visible in this analysis. However all ten samples have roughly equivalent total amounts of cyclized products, so their total stability is comparable.

### 5.5.2 Contribution of $\Delta$ Writhe

While the effect of  $\Delta$ Twist in this experiment is clear, the effect of  $\Delta$ Writhe is subtle. The modeled structure depicted in the chapter on design suggested that LZD73 would have a crossover angle that was near  $90^\circ$  while LZD87 would be closer to  $180^\circ$ . While this assay was not able to quantitate the absolute amount of writhe induced by the looping protein, a plot of the average Lk versus binding site separation (Figure 5.14) reveals that for all but the 443/453 set, the average is greater than the DNA-only value of 81.76.

If the only topological effect of looping were the  $\Delta$ Twist component, then, over the course of one helical repeat, the over-twisting and under-twisting events should ultimately average out. This is not to say that every binding site separation should have an  $Lk_{ave}$  equal to the DNA-only value, but rather that the average of all samples within a helical repeat should coincide with the DNA-only value. As seen in Figure 5.14, this is not the case. In fact, each repeat has only one sample with a  $Lk_{ave}$  value less than the DNA-only value and three samples that have a greater  $Lk_{ave}$ . We ascribe the overall positive change in Lk to be due to positive writhe induced by looping

If the untwisting of the binding event, suggested in Hockings et. al, is taken into account there should actually be an overall decrease in the  $Lk_{ave}$ . This could mean that the total  $\Delta$ Writhe component to this loop is actually greater than the values depicted in Figure 5.14. Additionally, the difference between the two protein samples indicated that there is a difference in writhe between the two. However, to

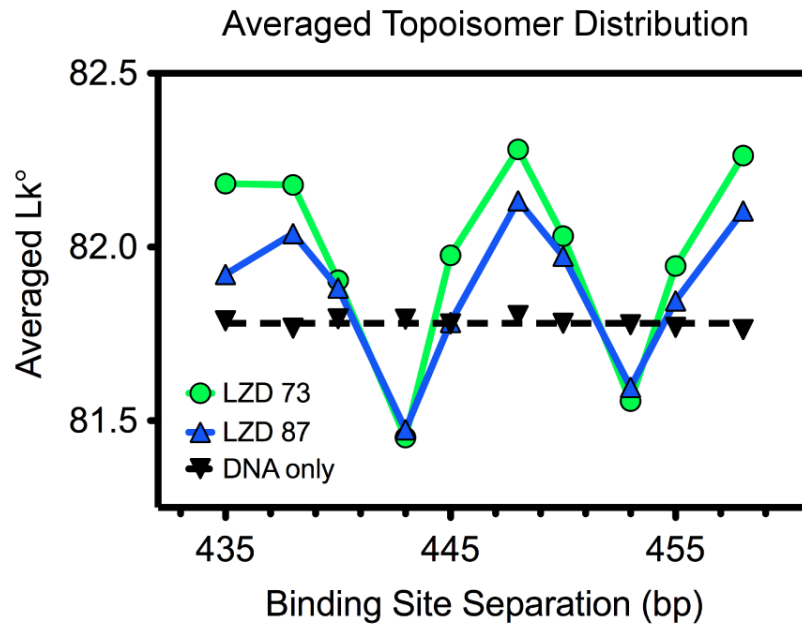


Figure 5.14 Weighted average topoisomer distributions were calculated for each Vx fragment. The black horizontal line is the DNA-only samples. The positive displacement from the DNA-only average and the difference between LZD73 and LZD83 are likely due to differences in the writhe induced by the two loops.

quantify this writhe and provide a measurement of the angle between bound DNA strands additional experiments will be necessary. One possible such experiment would to use FRET to measure the distance between bound DNA stands, which could provide the local information necessary to calculate the spatial positions.

## Chapter 6: Conclusions and Future Directions



## 6.1 *Demonstration of an Artificial DNA Looping Protein*

Following the example set by Bellomy et al. (1988) and Muller et al. (1996) our cyclization assays provided evidence of looping by demonstrating a dependency of results on the helical repeat of DNA. While we were clearly capable of forming loops, there is a surprising gap between the lower length limit established by our assay (310 bp) and that observed with natural proteins. Specifically, LacI can form a loop of 92 bp in its natural *lac* operon setting and has also been shown to form loops as small as 58 bp in an engineered system. This gap of nearly 250 bp is likely the result of the inherent flexibility in the LacI protein, allowing it to provide stability to the very short loop by conforming to the DNA's most energetically relaxed form. LZD proteins, in contrast, were engineered to avoid this flexibility and therefore any structural strain required to form a loop is isolated to the DNA molecule. This was a major goal of this project and further work will be needed to exploit LZD's potential as a rigid artificial DNA looping protein.

## 6.2 *Further Characterization of LZD proteins*

### 6.2.1 Identifying the Minimal Looping Length

The successful engineering of our protein opens up several new avenues of exploration. Initially, further characterization of the looping limits should be investigated. While the results from the Vx(153-448) cyclization work in section 5.3 identified 310 bp as the shortest length loop to stably form, our results cannot be used to support the claim that loops 309 bp or 308 bp would not form. Therefore a more

thorough analysis of loop formation between 200 bp and 300 bp is warranted. The use of the periodic binding site separation approach would be well suited as its results are easily interpreted. This would involve significant cloning to produce the plasmid templates required, making an alternative method of assembly appealing. A faster approach may be to utilize a modular assembly of the DNA using a combination PCR products and covalently linking them together. It is conceivable that 50 variable length looping fragments with 200 to 300 bp binding site separation could be assembled this way in less time than it took to clone the original ten plasmid templates for Vx(435-458) fragments. The determination of an explicit lower limit would be a tremendous asset in understanding the energy behind DNA looping.

#### 6.2.2 Measurement of the Relative DNA Binding Angle Using FRET

An additional level of binding site angle characterization could be achieved using a FRET based approach to measure the difference between LZD73 and LZD87 binding orientations. As seen in Figure 6.1, the placement of FRET donor and acceptor molecules on the ends of DNA situated near the CREB and Inv-2 site DNA would allow for measurement of their separation. The relative difference between LZD73 and LZD87 could be quantitated to provide a direct measurement of the binding angle. This could be very useful in putting these proteins to use in assembling protein:DNA nanostructures as the ability to rotate the bound DNA along the coiled-coil axis provides an additional dimension to the engineering platform.

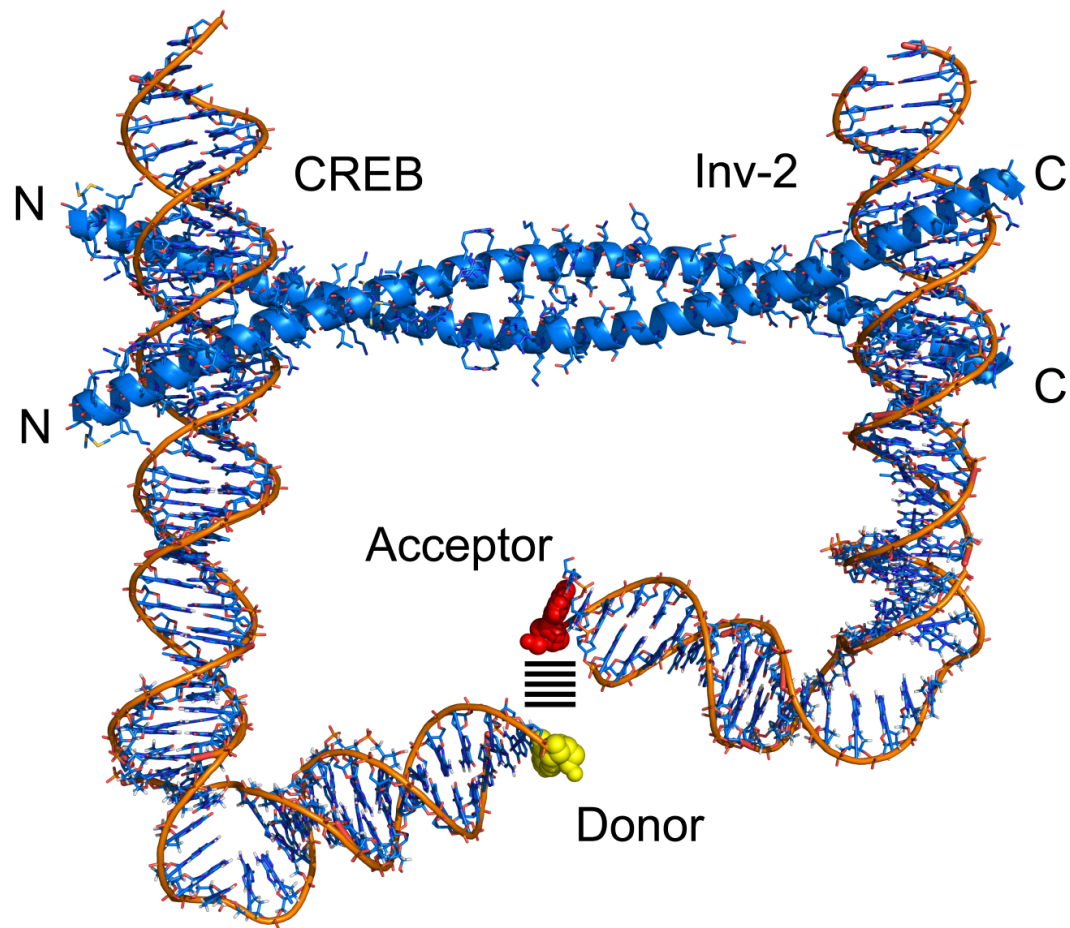


Figure 6.1 FRET based measurement of binding site orientation. DNA bulges (PDB:1JRV) were used to angle the DNA such that donor and acceptor are closer to each other than if simple places immediately next to the CREB and Inv-2 sites. This would make FRET measurement possible for a loop if the two binding sites were on the same DNA fragment, having both the CREB and Inv-2 sites.

### 6.3 *General Future Directions*

#### 6.3.1 The Creation of Topological Domains using LZD proteins

Recent work by Fenfei Leng demonstrated that LacI and GalR-mediated loops were capable of forming topologically isolated domains in plasmids DNA (Leng et

al., 2011). This was shown individually relaxing two asymmetrically sized loops formed in supercoiled plasmid by LacI binding. The LacI induced looping prevented the relaxation of one loop from affecting the supercoiling of the second loop thus

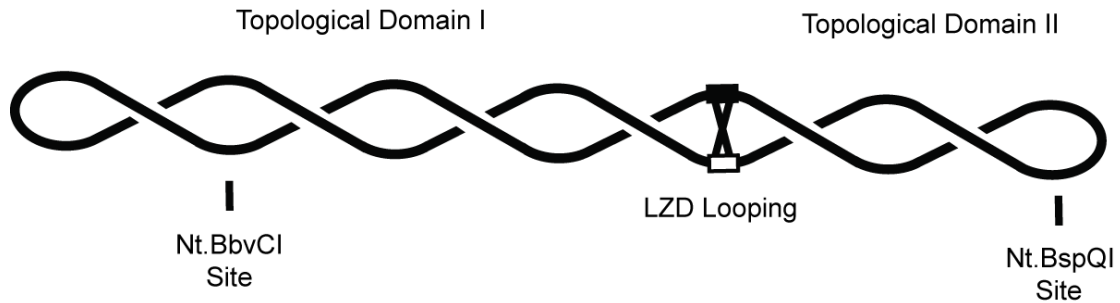


Figure 6.2 Topological domains formed by looping with LXD protein. A nicking enzyme recognition site is located in each of the two looped domains. Individual nicking should result in different amounts of plasmid relaxation because of the asymmetric size of the loops.

establishing topological isolation. This approach could be readily used to investigate LXD looping. The main benefit to this approach would be an ability to work in a more physiological buffer (as opposed to the T4 DNA ligase reaction's required low ionic strength). Figure 6.2 depicts how this reaction could be performed. Nt.BbvCI and Nt.BspQI are both site-specific DNA nicking enzymes. Following relaxation with these enzymes, the nicks would be enzymatically repaired with a ligase (presumably not T4 DNA ligase) and the plasmid resolved by agarose electrophoresis.

### 6.3.2 Protein:DNA Nanostructures in Two and Three Dimensions

A second major undertaking for LXD would involve its use in the assembly of protein:DNA nanostructures. Binding specificity and rotation of the N and C terminal binding sites through LXD73-87 14 aa coiled-coil extension, would allow for

the creation of highly ordered 2 and 3 dimensional nano-structures. The Kahn lab has been working to construct DNA triangle and square nanostructures within which LZD specific binding sites could be incorporated to direct protein binding in plane or out of plane with the DNA molecule. By rotating the binding site, though deletion and/or additions of bp on either side of the site, it is conceivable that protein:DNA cubes or triangular pyramids could be assembled and visualized by Atomic Force Microscopy (AFM).

### 6.3.3 Introducing a Flexible Hinge into LZD

To provide further proof that protein flexibility plays a large role in stabilizing short DNA loops, it would be valuable to introduce flexibility into our looping proteins. The most obvious way to do this would be to incorporate a linker sequence in the middle portion of the protein that would consist of random coil. This could be achieved by the use of prolines and glycines in the linker. Prolines are known to break  $\alpha$ -helices because of their inability to form appropriate hydrogen bonds and glycines allow the least steric hindrance and have greater entropy in the unfolded state. Their incorporation into the LZD peptides would create a hinge region that would allow the protein to bend and twist relative to the two binding sites. Figure 6.3 depicts how a hinge region within the structure may affect protein flexibility. If this new protein could be used to form loops significantly shorter than LZD73 or LZD87, it would be solid evidence that the flexibility of DNA looping proteins is an essential component of loop stability.

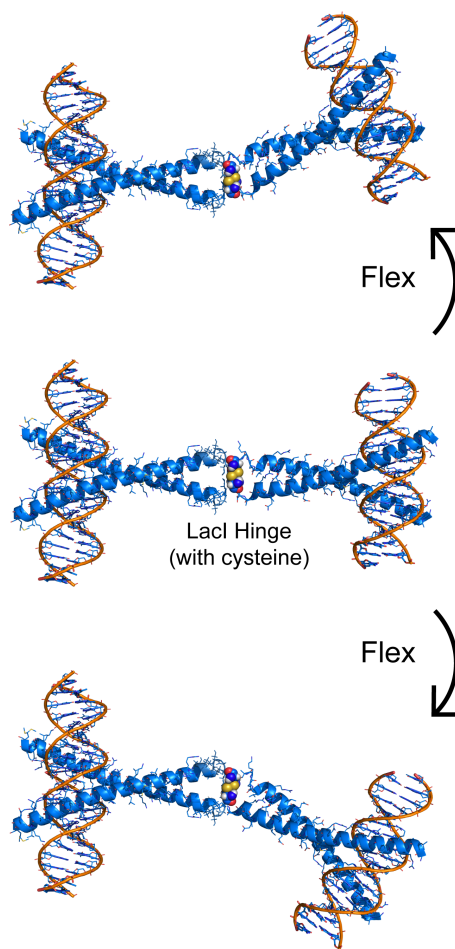


Figure 6.3 LZD Flex design incorporating a proline and glycine rich spacer to confer flexibility. Shown with an optional cysteine disulfide bridge to provide enhance stability of the protein dimer.

## 6.4 Conclusions

DNA loop formation, in nature, is mediated by a variety of looping proteins. Among those studied, it has been observed that the protein is inherently flexible and can therefore assume different conformations while forming a stable loop. It has been suggested that this protein flexibility adds to the stability of the loop and, based on the current model of DNA flexibility, allows for the formation of surprisingly small DNA

loops *in vivo*. To investigate whether such small loops were forming as a result of the protein's flexibility or if DNA is actually more flexible at shorter length than believed, we designed and synthesized a series of DNA looping proteins that were inherently rigid. This work represents the first successful effort in using DNA looping proteins in this manner.

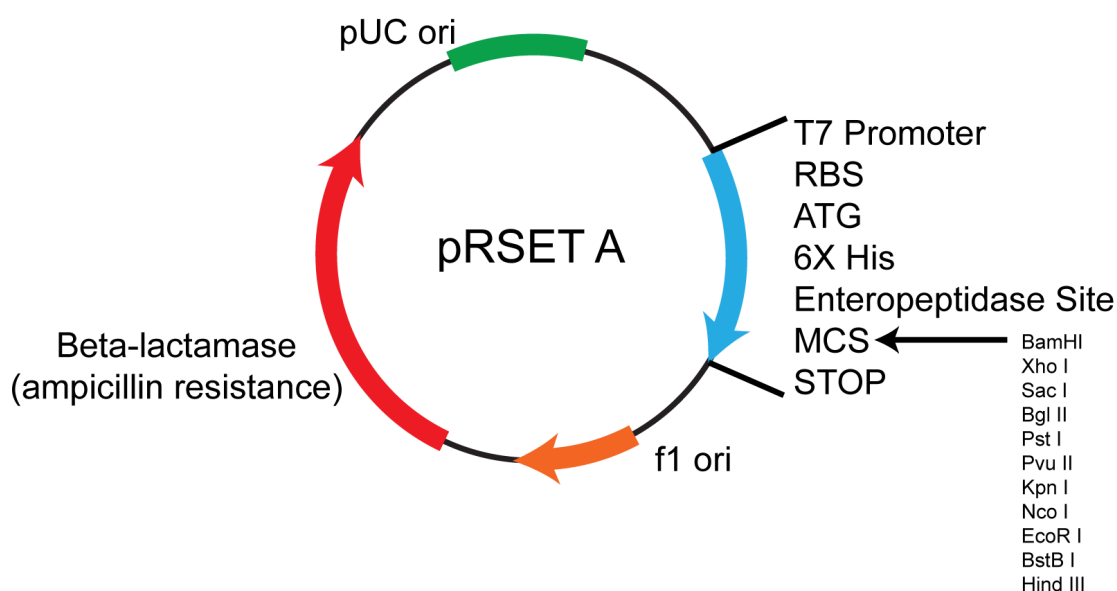
DNA cyclization experiments were used to characterize the geometry of the loops formed by capturing different DNA topoisomers that were a result of the topological strain induced on the DNA by the looping protein. Our results showed that loops smaller than 310 bp were difficult to form, but above 310 bp formed easily and were quite stable. This is evidence that hypothesis of DNA being inherently more flexible than previously believed is not true.

Furthermore, our artificial DNA looping proteins represent the first application of an artificial looping protein to manipulate DNA topology. These engineered proteins will be put to use in assembling protein:DNA nanostructures that rely on site-specific interactions. The demonstration of the coiled-coiled extension between LZD73 and LZD87 having an affect on the binding geometry means that, with simple protein engineering, this platform can be used to highly orchestrate the assembly of three-dimensional structures. This project represents the first step in harnessing this unique approach to engineering protein:DNA nanostructures.

## Appendix 1 - Sequences of Relevant DNA Constructs



All proteins were cloned into the ORF of pRSETA using BamHI and EcoRI cohesive ends and the corresponding insertion points in the multiple cloning site within the plasmid. The figure below is a plasmid map of pRSETA indicating the pUC origin of replication, the  $\beta$ -lactamase gene (which provides resistance to ampicillin), and the open reading frame (ORF) location (which is under the control of a T7 promoter).



Appendix 1 Plasmid map of pRSETA. Genes for the designed proteins were cloned into the MCS region of the expression vector (blue) using the BamHI and EcoRI sites. pUC origin (green) confers a high-copy number of this plasmid within *E. coli*, while ampicillin resistance is afforded by the presence of a  $\beta$ -lactamase gene (red)

#### Cloned Proteins: ORF and Translated Sequences

BamHI (GGATCC) and EcoRI (GAATTC) sites used for cloning have been underlined.

4HAR

ORF

ATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAG  
CAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCGATGGGGATCCGACCCA  
GCGGCACTGAAACGTGCACGCAACACCGAAGCTGCACGTCGTTCCCGTGCTCGTAAA  
CTGCAGCGTATGAAACAACCTGGAAGACAAAGTTGAAGAGCTGCTGTCCAAGAACTAC  
CACCTGGAAAACGAAGTTGCTCGTCTGAAAAAACTGGTTGCGCGCCTGGCGCGTCAA  
GTACGTGCTCTGGCTGATTCTCTGATGCAGCTGGCTCGCCAGGTTTCCCGTCTGGCA  
GACTCCCTGTAGTAGAATTC

Translated Sequence

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAALKRARNTAARRSRARK  
LQRMKQLEDKVEELLSKNYHLENEVARLKKLVARLARQVRALADSLMQLARQVSRLA  
DSL\*

4HEE

ORF

ATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAG  
CAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCGATGGGGATCCGACCCA  
GCGGCACTGAAACGTGCACGCAACACCGAAGCTGCACGTCGTTCCCGTGCTCGTAAA  
CTGCAGCGTATGAAACAACCTGGAAGACAAAGTTGAAGAGCTGCTGTCCAAGAACTAC  
CACCTGGAAAACGAAGTTGCTCGTCTGAAAAAACTGGTTGAGGAGCTGGCGCGTCAA  
GTACGTGCTCTGGCTGATTCTCTGATGCAGCTGGCTCGCCAGGTTTCCCGTCTGGCA  
GACTCCCTGTAGTAGAATTC

Translated Sequence

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSDPAALKRARNTAARRSRARK  
LQRMKQLEDKVEELLSKNYHLENEVARLKKLVEELARQVRALADSLMQLARQVSRLA  
DSL\*

LZAR

ORF

ATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAG  
CAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCGATGGGGATCCGACCCA  
GCGGCACTGAAACGTGCACGCAACACCGAAGCTGCACGTCGTTCCCGTGCTCGTAAA  
CTGCAGCGTATGAAACAACCTGGAAGACAAAGTTGAAGAGCTGCTGTCCAAGAACTAC  
CACCTGGAAAACGAAGTTGCTCGTCTGAAAAAACTGGTTGCGCGCCTGAAGAAGCTG  
GTACGTGCTCTGGCTGATTCTCTGATGCAGCTGGCTCGCCAGGTTTCCCGTCTGGAA  
TCCGGTCAGTAGTAGAATTC

Translated Sequence

MRGSHHHHHHGMASMTGGQOMGRDLYDDDDKDRWGSDPAALKRARNTAAARRSRARK  
LQRMKQLEDKVEELLSKNYHLENEVARLKKLVARLKKLVRALADSLMQLARQVSRLE  
SGQ\*

LZEE

ORF

ATGCGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAG  
CAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCGATGGGGATCCGACCCA  
GCGGCACTGAAACGTGCACGCAACACCGAAGCTGCACGTCGTTCCCGTGCTCGTAAA  
CTGCAGCGTATGAAACAACCTGGAAGACAAAGTTGAAGAGCTGCTGTCCAAGAACTAC  
CACCTGGAAAACGAAGTTGCTCGTCTGAAAAAACTGGTTGAGGAGCTGCTGTCCAAA  
GTACGTGCTCTGGCTGATTCTCTGATGCAGCTGGCTCGCCAGGTTTCCCGTCTGGAA  
TCCGGTCAGTAGTAGAATTC

Translated Sequence

MRGSHHHHHHGMASMTGGQOMGRDLYDDDDKDRWGSDPAALKRARNTAAARRSRARK  
LQRMKQLEDKVEELLSKNYHLENEVARLKKLVEELLSKVRALADSLMQLARQVSRLE  
SGQ\*

GCN4 bZip control peptide aa sequence (provided by Jim  
Maher)

MRGSHHHHHHRSMGRDPAALKRARNTAAARRSRARKLQRMKQLEDKVEELLSKNYHL  
ENEVARLKKL

*reverse*GCN4

ORF

ATGCGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAG  
CAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCGATGGGGATCCATGCAA  
CGAATGAAGCAGCTGGAAGACAAGGTGGAGGAAGCTGCTGAGCAAGAACTACCACCTG  
GAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGAAGTTACAGCGG  
GTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGAAAGGCTGCT  
CTGAAGGGATAGTAAGAATTC

Translated Sequence

MRGSHHHHHHGMASMTGGQOMGRDLYDDDDKDRWGSMDQRMKQLEDKVEELLSKNYHL  
ENEVARLKKLVGELQKLQRVKRARNTAAARRSRARKAALKG\*\*

LZD73

ORF

ATGCGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAG  
CAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCGATGGGGATCCGATCCA  
GCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAG  
CTGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAAGTAC  
CACCTGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGAAGTTA  
CAGCGGGTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGAAAG  
GCTGCTCTGAAGGGATAGTAAGAATTC

Translated Sequence

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGS DPAALKRARNT EAARRSRARK  
LQRMKQLEDKVEELLSKNYHLENEVARLKKLVGELQKLQRVKRARNT EAARRSRARK  
AALKG\*\*

LZD87

ORF

ATGCGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAG  
CAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCGATGGGGATCCGATCCA  
GCTGCTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAG  
CTGCAACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAAGTAC  
CACCTGGAGAACGAAGTTGCGCGCCTGAAAAGCTGGTGGAGAAGTCTGAGCAAAA  
GTGCGTGCGCTGGCGGATTCTCTGGGCGAACTGCAGAAGTTACAGCGGGTGAAGCGA  
GCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGAAAGGCTGCTCTGAAGGGA  
TAGTAAGAATTC

Translated Sequence

MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGS DPAALKRARNT EA  
RRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVEELLSKVRLA  
DSL GELQKLQRVKRARNT EAARRSRARKAALKG\*\*

EMSA oligos– Binding Sites **HIGHLIGHTED**

Assembled By PCR:

177 bp InvB Loop

GGGATCCCTGACGGCGCGC**ATGACGTCAT**GGAATTCGAAGCTTGATCCGGCTGCTAA  
CAAAGCCC GAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATA

ACCCCTTGGCTGCACATGACGTCATGCGCGGGATCCGAATTCTCCGGTCTGGCGTA  
ATAGCG

Ix1 used with GCN4 bZip *reverse*GCN4 and LZD73

GAACAACACTCAACCCTATCTCGCGTCATATGACCAAGCTGAATTCGCGCGCTGACC  
TCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATAT  
GTATCCGCTCATGAGACAATAACCCCTCGAGTTGCAGCTTTT

Annealed using IDT ssDNA oligos

CREB 58 bp A

CGTCGACGAGGCCGAGCTCTGAGGATGACGTCATAAGCAGCTGGAAGACTTCTGCAG  
G

CREB 58 bp B

CCTGCAGAAGTCTTCCAGCTGCTTATGACGTCATCCTCAGAGCTCGGCCTCGTCGAC  
G

Inv-2 30 bp A

GATCTTAACACTGTCATATGACTAGAGCGG

Inv-2 30 bp B

CCGCTCTAGTCATATGACAGTGTTAAGATC

DNA products for Dimerization Experiments

Binding Sites **HIGHLIGHTED**

XhoI Sites underlined

Ix1

GAACAACACTCAACCCTATCTCGCGTCATATGACCAAGCTGAATTCGCGCGCTGACC  
TCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATAT  
GTATCCGCTCATGAGACAATAACCCCTCGAGTTGCAGCTTTT

Ix2

GAACAACACTCAACCCTATCTCGCGTCATATGACCAAGCTGAATTCGCGCGCTGACC  
TCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATAT  
GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAA  
GAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTTCGCGCATTTTG  
CCTTCCTGTTTCTCGAGTTGCAGCTTTT

Ix3

GAACAACACTCAACCCTATCTCGCGTCATATGACCAAGCTGAATTCGCGCGCTGACC  
TCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATAT  
GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAA  
GAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTTCGCGCATTTTG  
CCTTCCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCA  
GTTGGGTGCACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAGATCCTCTC  
GAGTTGCAGCTTTT

#### DNA Products for Cyclization Experiments

Vx(153)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCGCTGC  
TCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGCA  
ACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACCT  
GGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCATG  
CGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATCG  
CCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGC  
ATTAAGCGCGGCGGGTGTGGTGGTTACGTCATATGACCAAGCTGAATTCGCGCGCT  
GACCTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAA  
ATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAA  
GGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTTCGCGCAT  
TTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(202)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTAGCGCCCGCTCCTTTCGGTCATATGACCAAGCTGAATTCGCGCGCTGACCTCG  
GAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATC  
CGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTA

TGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTC  
CTGTTTCTCGAGTTGCAGCTTTT

Vx(254)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAACGTCCGTCAATATGACCAAGCTGAATTCGCGCGCTGACCTCGGAAAT  
GTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTC  
ATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGT  
ATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTT  
TCTCGAGTTGCAGCTTTT

Vx(310)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCTGTCAATATGACCAAGCTGAATTCGCGCGCTGACCTCGGAAATG  
TGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCA  
TGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTA  
TTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTT  
CTCGAGTTGCAGCTTTT

Vx(376)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG

CCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTTCGCCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC  
CCTGATAGACGGTTTTTCGCCCTGTCATATGACCAAGCTGAATTCGCGCGCTGACC  
TCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGT  
ATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA  
GTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTTTTTCGCGCATTTTGCC  
TTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(448)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCGACGA  
TAAGGATCGATGGGGATCCGATCCAGCTGCTCTGAAGCGAGCTCGGAACACTGAAGC  
TGCTCGACGGAGCCGAGCTCGGAAGCTGCAACGAATGAAGCAGCTGGAAGACAAGGT  
GGAGGAACTGCTGAGCAAGAACTACCACCTGGAGAACGAAGTTGCGCGCCTGAAGAA  
GCTGGTGGGTGAACTGCAGATGACGTCATGCGCGCGGATCCGAATTCTCCGGATCTG  
GCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAA  
TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGCGGGTGTGGTGGTTAC  
GCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTT  
CCCTTCTTTCTCGCCACGTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCT  
CCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTA  
GGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTTCGCCCTTTGAC  
GTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAA  
CCCTATCTCGCGTCATATGACCAAGCTGAATTCGCGCGCTGACCTCGGGGAAATGTG  
CGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATG  
AGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATT  
CAACATTTCCGTGTCGCCCTTATTCCTTTTTTTCGCGCATTTTGCCTTCCTGTTTCT  
CGAGTTGCAGCTTTT

Vx(448)100

AATGTGTGCCTGGCGATCTCGAGTGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAG  
CAAGAACTACCACCTGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACT  
GCAGATGACGTCATGCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGA  
GGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGC  
GCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGC  
TACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGC  
CACGTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCG  
ATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACG  
TAGTGGGCCATCGCCCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTT  
CTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGCGTCA  
TATGACCAAGCTTGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC  
TGCCACCGCTGAGCAATAACTAGCATAACCCCTTTTTCAGCTTTTCTCGAGTCAAGA  
CCCGTTTAGAGGCCCC



The additional 9 PCR constructs made for 435–458 Phase Cyclization

Vx(435)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAAGTCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAAGTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTATCGCCCGCTCCTTTCGCTTTCCTTCCCTTCCTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCCAAAAAAGTGAATAGGGTGTGGTTCACGTAGTGGGCCATCGC  
CCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC  
TCTTGTTCCAACTGGAACAACACTCGTCATATGACAACCCTATAAGCTTAAGCTGA  
ATTGCGCGCTGACCTCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA  
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT  
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTGCGCCCTTATTCCTTTT  
TTGCGGCATTTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(438)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAAGTCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAAGTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTATCGCCCGCTCCTTTCGCTTTCCTTCCCTTCCTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCCAAAAAAGTGAATAGGGTGTGGTTCACGTAGTGGGCCATCGC  
CCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC  
TCTTGTTCCAACTGGAACAACACTCAACGTCATATGACCCTATAAGCTTAAGCTGA  
ATTGCGCGCTGACCTCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA  
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT  
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTGCGCCCTTATTCCTTTT  
TTGCGGCATTTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(440)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC

AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTATCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC  
CCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC  
TCTTGTTCCAACTGGAACAACACTCAACCCGTCATATGACTATAAGCTTAAGCTGA  
ATTGCGCGCTGACCTCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA  
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT  
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTT  
TTGCGGCATTTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(443)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTATCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC  
CCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC  
TCTTGTTCCAACTGGAACAACACTCAACCCCTATGTCATATGACAAGCTTAAGCTGA  
ATTGCGCGCTGACCTCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA  
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT  
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTT  
TTGCGGCATTTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(445)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTATCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC

CCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC  
TCTTGTTCCAACTGGAACAACACTCAACCCTATAAGTCATATGACGCTTAAGCTGA  
ATTCGCGCGCTGACCTCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA  
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT  
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTT  
TTGCGGCATTTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(450)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTATCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC  
CCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC  
TCTTGTTCCAACTGGAACAACACTCAACCCTATAAGCTTAGTCATATGACAGCTGA  
ATTCGCGCGCTGACCTCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA  
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT  
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTT  
TTGCGGCATTTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(453)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAACTACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTATCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC  
CCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC  
TCTTGTTCCAACTGGAACAACACTCAACCCTATAAGCTTAAGCGTCATATGACTGA  
ATTCGCGCGCTGACCTCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA  
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT  
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTT  
TTGCGGCATTTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(455)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAATAACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTATCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCCAAAAAAGTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC  
CCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC  
TCTTGTTCCAACTGGAACAACACTCAACCCTATAAGCTTAAGCTGGTCATATGACA  
ATTCGCGCGCTGACCTCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA  
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT  
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTGCGCCCTTATTCCTTTT  
TTGCGGCATTTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

Vx(458)200

AATGTGTGCCTGGCGATCTCGAGGAGGATAAGGATCGATGGGGATCCGATCCAGCTG  
CTCTGAAGCGAGCTCGGAACACTGAAGCTGCTCGACGGAGCCGAGCTCGGAAGCTGC  
AACGAATGAAGCAGCTGGAAGACAAGGTGGAGGAACTGCTGAGCAAGAATAACCACC  
TGGAGAACGAAGTTGCGCGCCTGAAGAAGCTGGTGGGTGAACTGCAGATGACGTCAT  
GCGCGCGGATCCGAATTCTCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATC  
GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCG  
CATTAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCG  
CCCTATCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTGCGCGGCT  
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTAC  
GGCACCTCGACCCCCAAAAAAGTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC  
CCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC  
TCTTGTTCCAACTGGAACAACACTCAACCCTATAAGCTTAAGCTGAATGTCATATG  
ACTCGCGCGCTGACCTCGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA  
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAAT  
ATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTGCGCCCTTATTCCTTTT  
TTGCGGCATTTTGCCTTCCTGTTTCTCGAGTTGCAGCTTTT

## Appendix 2: Additional Protein Purification Procedures

Method Protocol for the AKTA Unicorn program that runs the HiTrap Chelating Column ( $\text{Co}^{2+}$ ) for LZD and 4HB mutants. This program reflects the optimized protocol which implements a 0-70 % gradient with elution buffer over a 25 CV (25 mL) volume. Early purification work with the 4HB mutants involved a 0-100 % gradient over 40 CV. The current approach provides superior protein resolution and requires less buffer consumption.

Buffers used for LZAR, 4HAR, and 4HEE:

Equilibration – 10 mM MES pH 6.0, 6 M Guanidine, 0.5 M NaCl, 20 mM Imidazole.

Elution – 10 mM MES pH 6.0, 6 M Guanidine, 0.5 M NaCl, 0.4 M Imidazole.

Buffers used for LZEE, LZD73, and LZD87:

Equilibration – 10 mM MES pH 6.0, 0.5 M NaCl, 20 mM Imidazole.

Elution – 10 mM MES pH 6.0, 0.5 M NaCl, 0.4 M Imidazole.

Method: c:\UNICORN\Local\fil\Dan AKTA\method\HiTrap Chelating 1 mL Gradient Elution.m01

Main method:

☐ Main

0.00 Base CV 0.96 {ml} HiTrap\_Chelating\_HP\_1\_ml

☐ 0.00 Block Starting\_Conditions

Starting\_Conditions

0.00 Base SameAsMain

0.00 AutozeroUV

0.00 Alarm\_Pressure Enabled 0.3 {MPa} 0.000 {MPa}

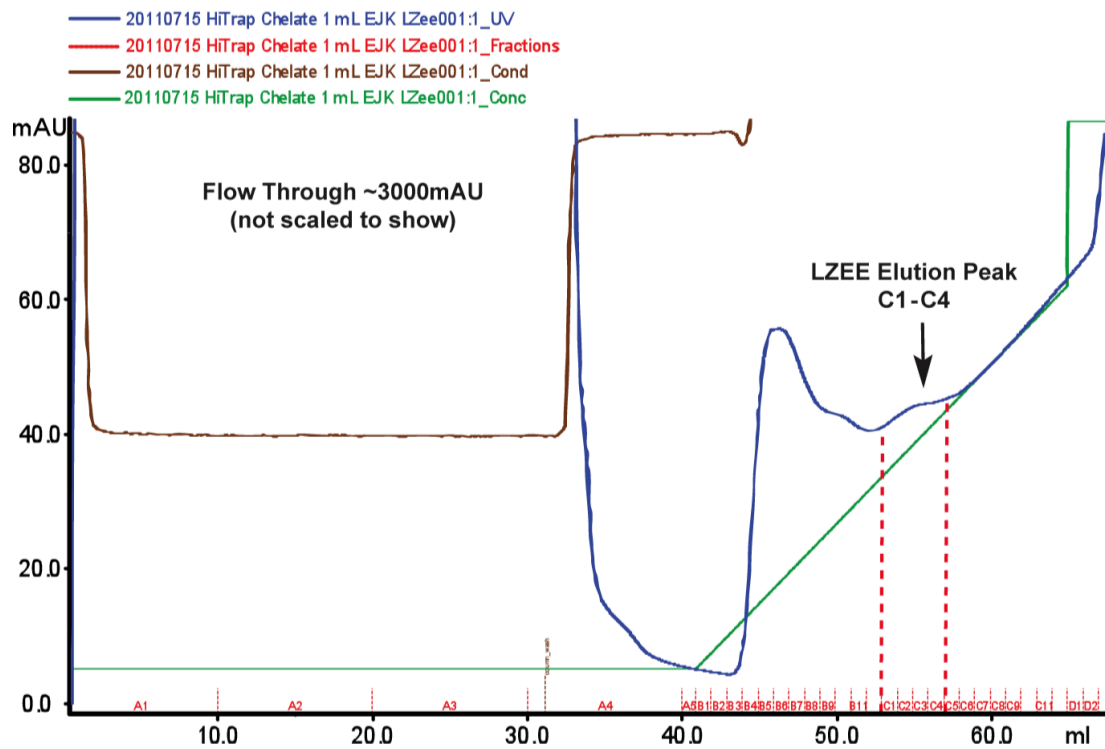
0.00 Flow 1.00 {ml/min}

0.00 AveragingTimeUV 1.30

```

0.00 ColumnPosition Position2
0.00 BufferValveA A1
0.00 InjectionValve Load
0.00 End_Block
☐ 0.00 Block Equilibration
    Equilibration
    0.00 Base SameAsMain
    0.00 Flow 1.0 {ml/min}
    0.00 ColumnPosition Position2
    0.00 BufferValveA A1
    8.00 AutozeroUV
    8.00 End_Block
☐ 0.00 Block Load
    Load
    0.00 Base SameAsMain
    0.00 InjectionValve Inject
    0.00 Fractionation 18 mm 8.0 {ml} FirstTube Volume
    0.00 Flow 1.0 {ml/min}
    0.00 Set_Mark "Sample Load"
    (15.00)#Load_Volume InjectionValve Load
    15.00 Set_Mark "Column Wash"
    15.00 End_Block
☐ 0.00 Block Wash
    Wash
    0.00 Base SameAsMain
    0.00 BufferValveA A1
    0.00 Flow 1.0 {ml/min}
    5.00 FractionationStop
    5.00 End_Block
☐ 0.00 Block Elution
    Elution
    0.00 Base SameAsMain
    0.00 Fractionation 18 mm 1.0 {ml} TubeNumber[B.1] Volume
    0.00 Gradient 70 {%B} (25)#Gradient_Length_CV {base}
    30.00 FractionationStop
    32.00 End_Block
0.00 End_Method

```



Purification LZEE in a single step HiTrap Chelating gradient run

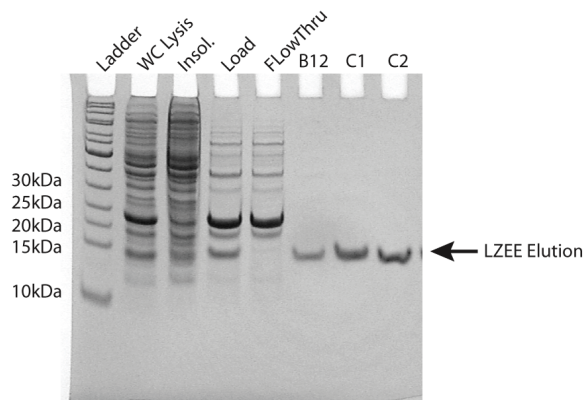
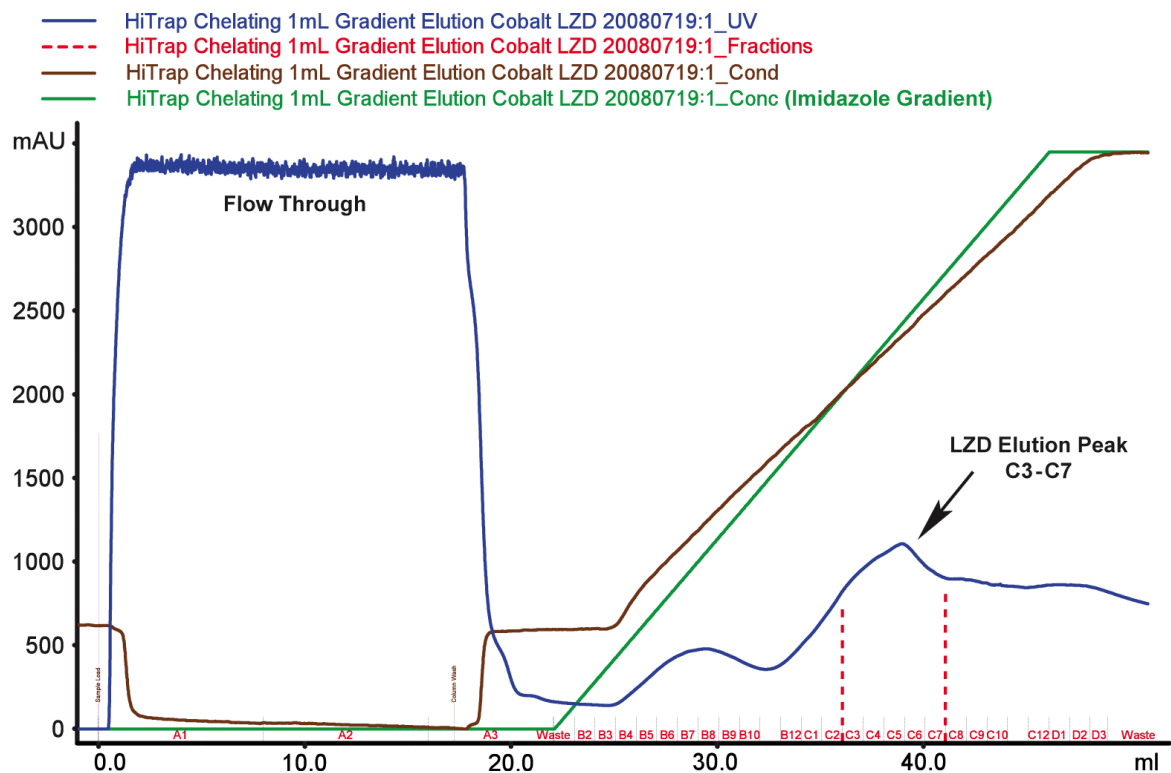


Figure 2.9 (repeated) Analysis of LZEE purification steps using HiTrap chelating affinity column purification with  $\text{Co}^{2+}$  metal. The gel analysis shows the whole cell lysis (WC), the insoluble pellet, the soluble load, the flow through, and then selected elution fractions B12-C2, corresponding to approximately [240-260 mM] imidazole, showing the target LZEE protein. For the chromatogram for this run, see appendix 2.





Purification LZD73 in a single step HiTrap Chelating gradient run

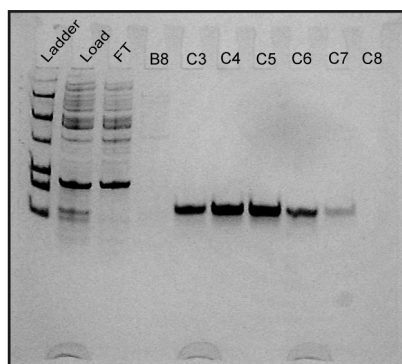


Figure 2.10 (repeated) Analysis of LZD 73 purification steps using HiTrap chelating affinity column purification with  $\text{Co}^{2+}$  metal. The gel analysis shows the load, the flow through (FT) and selected elution fractions B8 and C3-C8, corresponding to approximately [240-270 mM] imidazole.

## Bibliography

- Abate, C., Luk, D., Gentz, R., Rauscher, F. J., & Curran, T. (1990). Expression and purification of the leucine zipper and DNA-binding domains of Fos and Jun: both Fos and Jun contact DNA directly. *Proceedings of the National Academy of Sciences of the United States of America*, 87(3), 1032–1036.
- Alberti, S., Oehler, S., Wilcken-Bergmann, von, B., & Müller-Hill, B. (1993). Genetic analysis of the leucine heptad repeats of Lac repressor: evidence for a 4-helical bundle. *The EMBO journal*, 12(8), 3227–3236.
- Alberti, S., Oehler, S., Wilcken-Bergmann, von, B., Krämer, H., & Müller-Hill, B. (1991). Dimer-to-tetramer assembly of Lac repressor involves a leucine heptad repeat. *The New biologist*, 3(1), 57–62.
- Amoutzias, G. D., Veron, A. S., Weiner, J., Robinson-Rechavi, M., Bornberg-Bauer, E., Oliver, S. G., & Robertson, D. L. (2007). One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. *Molecular biology and evolution*, 24(3), 827–835. doi:10.1093/molbev/msl211
- Becker, N. A., Kahn, J. D., & Maher, L. J. (2005). Bacterial repression loops require enhanced DNA flexibility. *Journal of molecular biology*, 349(4), 716–730. doi:10.1016/j.jmb.2005.04.035
- Bellomy, G. R., Mossing, M. C., & Record, M. T. (1988). Physical properties of DNA in vivo as probed by the length dependence of the lac operator looping process. *Mutation of Ser-50 and Cys-66 in Snapin Modulates Protein Structure and Stability*, 27(11), 3900–3906.
- Burkhard, P., Kammerer, R. A., Steinmetz, M. O., Bourenkov, G. P., & Aepli, U. (2000). The coiled-coil trigger site of the rod domain of cortexillin I unveils a distinct network of interhelical and intrahelical salt bridges. *Structure (London, England : 1993)*, 8(3), 223–230.
- Camerini-Otero, R. D., & Felsenfeld, G. (1977). Supercoiling energy and nucleosome formation: the role of the arginine-rich histone kernel. *Nucleic acids research*, 4(5), 1159–1181.
- Chae, S. Y., Kim, H. J., Lee, M. S., Jang, Y. L., Lee, Y., Lee, S. H., Lee, K., et al. (2011). Energy-independent intracellular gene delivery mediated by polymeric biomimetics of cell-penetrating peptides. *Macromolecular bioscience*, 11(9), 1169–1174. doi:10.1002/mabi.201100088
- Chan, I.-S., Fedorova, A. V., & Shin, J. A. (2007). The GCN4 bZIP targets noncognate gene regulatory sequences: quantitative investigation of binding at full and half sites. *Biochemistry*, 46(6), 1663–1671. doi:10.1021/bi0617613

- Choder, M., & Aloni, Y. (1988). RNA polymerase II allows unwinding and rewinding of the DNA and thus maintains a constant length of the transcription bubble. *The Journal of biological chemistry*, 263(26), 12994–13002.
- Choe, S., & Sun, S. X. (2005). The elasticity of alpha-helices. *The Journal of chemical physics*, 122(24), 244912. doi:10.1063/1.1940048
- Choe, S., & Sun, S. X. (2007). Bending elasticity of anti-parallel beta-sheets. *Biophysical journal*, 92(4), 1204–1214. doi:10.1529/biophysj.106.095786
- Ciani, B., Bjelic, S., Honnappa, S., Jawhari, H., Jaussi, R., Payapilly, A., Jowitt, T., et al. (2010). Molecular basis of coiled-coil oligomerization-state specificity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46), 19850–19855. doi:10.1073/pnas.1008502107
- Cloutier, T. E., & Widom, J. (2004). Spontaneous sharp bending of double-stranded DNA. *Molecular cell*, 14(3), 355–362.
- Cloutier, T. E., & Widom, J. (2005). DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(10), 3645–3650. doi:10.1073/pnas.0409059102
- Coburn, J., Gibson, M., Bandalini, P. A., Laird, C., Mao, H.-Q., Moroni, L., Seliktar, D., et al. (2011). Biomimetics of the Extracellular Matrix: An Integrated Three-Dimensional Fiber-Hydrogel Composite for Cartilage Tissue Engineering. *Smart structures and systems*, 7(3), 213–222.
- Dalla-Favera, R., Gelmann, E. P., Martinotti, S., Franchini, G., Papas, T. S., Gallo, R. C., & Wong-Staal, F. (1982). Cloning and characterization of different human sequences related to the onc gene (v-myc) of avian myelocytomatosis virus (MC29). *Proceedings of the National Academy of Sciences of the United States of America*, 79(21), 6497–6501.
- Davison, L. J., Wallace, C., Cooper, J. D., Cope, N. F., Wilson, N. K., Smyth, D. J., Howson, J. M. M., et al. (2012). Long-range DNA looping and gene expression analyses identify DEXI as an autoimmune disease candidate gene. *Human molecular genetics*, 21(2), 322–333. doi:10.1093/hmg/ddr468
- Delbrück, M. (1954). ON THE REPLICATION OF DESOXYRIBONUCLEIC ACID (DNA). *Proceedings of the National Academy of Sciences of the United States of America*, 40(9), 783–788.
- Du, Q., Smith, C., Shiffeldrim, N., Vologodskaia, M., & Vologodskii, A. (2005). Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proceedings of the National Academy of Sciences of the United States of America*, 102(15), 5397–5402. doi:10.1073/pnas.0500983102
- Edelman, L. M., Cheong, R., & Kahn, J. D. (2003). Fluorescence resonance energy transfer over approximately 130 basepairs in hyperstable lac repressor-DNA loops. *Biophysical journal*, 84(2 Pt 1), 1131–1145. doi:10.1016/S0006-3495(03)74929-7
- Ellenberger, T. E., Brandl, C. J., Struhl, K., & Harrison, S. C. (1992). The GCN4 basic region

- leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell*, 71(7), 1223–1237.
- Emberly, E. G., Mukhopadhyay, R., Tang, C., & Wingreen, N. S. (2004). Flexibility of beta-sheets: principal component analysis of database protein structures. *Proteins*, 55(1), 91–98. doi:10.1002/prot.10618
- Faix, J., Steinmetz, M., Boves, H., Kammerer, R. A., Lottspeich, F., Mintert, U., Murphy, J., et al. (1996). Cortexillins, major determinants of cell shape and size, are actin-bundling proteins with a parallel coiled-coil tail. *Cell*, 86(4), 631–642.
- Ferré-D'Amaré, A. R., Pognonec, P., Roeder, R. G., & Burley, S. K. (1994). Structure and function of the b/HLH/Z domain of USF. *The EMBO journal*, 13(1), 180–189.
- Fisher, D. E., Parent, L. A., & Sharp, P. A. (1993). High affinity DNA-binding Myc analogs: recognition by an alpha helix. *Cell*, 72(3), 467–476.
- Friedman, A. M., Fischmann, T. O., & Steitz, T. A. (1995). Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science (New York, N.Y.)*, 268(5218), 1721–1727.
- Geanacopoulos, M., Vasmatazis, G., Zhurkin, V. B., & Adhya, S. (2001). Gal repressosome contains an antiparallel DNA loop. *Nature structural biology*, 8(5), 432–436. doi:10.1038/87595
- Geggier, S., & Vologodskii, A. (2010). Sequence dependence of DNA bending rigidity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(35), 15421–15426. doi:10.1073/pnas.1004809107
- Gellert, M., Mizuuchi, K., O'Dea, M. H., & Nash, H. A. (1976). DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 73(11), 3872–3876.
- Gerisch, G., Faix, J., Köhler, J., & Müller-Taubenberger, A. (2004). Actin-binding proteins required for reliable chromosome segregation in mitosis. *Cell motility and the cytoskeleton*, 57(1), 18–25. doi:10.1002/cm.10150
- Gilbert, W., & Maxam, A. (1973). The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences of the United States of America*, 70(12), 3581–3584.
- Gilbert, W., & Müller-Hill, B. (1966). Isolation of the lac repressor. *Proceedings of the National Academy of Sciences of the United States of America*, 56(6), 1891–1898.
- Haber, R., & Adhya, S. (1988). Interaction of spatially separated protein-DNA complexes for control of gene expression: operator conversions. *Proceedings of the National Academy of Sciences of the United States of America*, 85(24), 9683–9687.
- Haeusler, A. R., Goodson, K. A., Lillian, T. D., Wang, X., Goyal, S., Perkins, N. C., & Kahn, J. D. (2012). FRET studies of a landscape of Lac repressor-mediated DNA loops. *Nucleic acids research*. doi:10.1093/nar/gks019

- Hagerman, P. J. (1981). Investigation of the flexibility of DNA using transient electric birefringence. *Biopolymers*, 20(7), 1503–1535. doi:10.1002/bip.1981.360200710
- Hande, K. R. (1998). Clinical applications of anticancer drugs targeted to topoisomerase II. *Biochimica et biophysica acta*, 1400(1-3), 173–184.
- Hockings, S. C., Kahn, J. D., & Crothers, D. M. (1998). Characterization of the ATF/CREB site and its complex with GCN4. *Proceedings of the National Academy of Sciences of the United States of America*, 95(4), 1410–1415.
- Hollenbeck, J. J., Gurnon, D. G., Fazio, G. C., Carlson, J. J., & Oakley, M. G. (2001). A GCN4 variant with a C-terminal basic region binds to DNA with wild-type affinity. *Mutation of Ser-50 and Cys-66 in Snapin Modulates Protein Structure and Stability*, 40(46), 13833–13839.
- Hope, I. A., & Struhl, K. (1985). GCN4 protein, synthesized in vitro, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. *Cell*, 43(1), 177–188.
- Hope, I. A., & Struhl, K. (1987). GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *The EMBO journal*, 6(9), 2781–2784.
- Horowitz, D. S., & Wang, J. C. (1984). Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *Journal of molecular biology*, 173(1), 75–91.
- Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology*, 3, 318–356.
- Jacobson, H., Beckmann, C. O., & Stockmayer, W. H. (1950). Intramolecular Reaction in Polycondensations. II. Ring-Chain Equilibrium in Polydecamethylene Adipate. *The Journal of chemical physics*, 18(12), 1607. doi:10.1063/1.1747548
- Kahn, J. D., & Crothers, D. M. (1998). Measurement of the DNA bend angle induced by the catabolite activator protein using Monte Carlo simulation of cyclization kinetics. *Journal of molecular biology*, 276(1), 287–309. doi:10.1006/jmbi.1997.1515
- Kahn, J. D., Yun, E., & Crothers, D. M. (1994). Detection of localized DNA flexibility. *Nature*, 368(6467), 163–166. doi:10.1038/368163a0
- Kammerer, R. A., Schulthess, T., Landwehr, R., Lustig, A., Engel, J., Aebi, U., & Steinmetz, M. O. (1998). An autonomous folding unit mediates the assembly of two-stranded coiled coils. *Proceedings of the National Academy of Sciences of the United States of America*, 95(23), 13419–13424.
- Kavenoff, R., & Bowen, B. C. (1976). Electron microscopy of membrane-free folded chromosomes from Escherichia coli. *Chromosoma*, 59(2), 89–101.
- Keller, W., König, P., & Richmond, T. J. (1995). Crystal structure of a bZIP/DNA complex at 2.2 Å: determinants of DNA specific recognition. *Journal of molecular biology*, 254(4), 657–667. doi:10.1006/jmbi.1995.0645

- Kikuchi, A., & Asai, K. (1984). Reverse gyrase--a topoisomerase which introduces positive superhelical turns into DNA. *Nature*, 309(5970), 677–681.
- Koldin, B., Suckow, M., Seydel, A., Wilcken-Bergmann, von, B., & Müller-Hill, B. (1995). A comparison of the different DNA binding specificities of the bZip proteins C/EBP and GCN4. *Nucleic acids research*, 23(20), 4162–4169.
- Krämer, H., Niemöller, M., Amouyal, M., Revet, B., Wilcken-Bergmann, von, B., & Müller-Hill, B. (1987). lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *The EMBO journal*, 6(5), 1481–1491.
- Lamond, A. I. (1985). Supercoiling response of a bacterial tRNA gene. *The EMBO journal*, 4(2), 501–507.
- Landschulz, W. H., Johnson, P. F., & McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science (New York, N.Y.)*, 240(4860), 1759–1764.
- Lebel, R., McDuff, F.-O., Lavigne, P., & Grandbois, M. (2007). Direct visualization of the binding of c-Myc/Max heterodimeric b-HLH-LZ to E-box sequences on the hTERT promoter. *Mutation of Ser-50 and Cys-66 in Snapin Modulates Protein Structure and Stability*, 46(36), 10279–10286. doi:10.1021/bi700076m
- Leng, F., Chen, B., & Dunlap, D. D. (2011). Dividing a supercoiled DNA molecule into two independent topological domains. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), 19973–19978. doi:10.1073/pnas.1109854108
- Lewis, D. E., Geanakopoulos, M., & Adhya, S. (1999). Role of HU and DNA supercoiling in transcription repression: specialized nucleoprotein repression complex at gal promoters in Escherichia coli. *Molecular microbiology*, 31(2), 451–461.
- Lobell, R. B., & Schleif, R. F. (1991). AraC-DNA looping: orientation and distance-dependent loop breaking by the cyclic AMP receptor protein. *Journal of molecular biology*, 218(1), 45–54.
- Maxwell, A., & Lawson, D. M. (2003). The ATP-binding site of type II topoisomerases as a target for antibacterial drugs. *Current topics in medicinal chemistry*, 3(3), 283–303.
- McDonald, R. J., Dragan, A. I., Kirk, W. R., Neff, K. L., Privalov, P. L., & Maher, L. J. (2007). DNA bending by charged peptides: electrophoretic and spectroscopic analyses. *Biochemistry*, 46(9), 2306–2316. doi:10.1021/bi061921a
- Mehta, R. A., & Kahn, J. D. (1999). Designed hyperstable Lac repressor.DNA loop topologies suggest alternative loop geometries. *Journal of molecular biology*, 294(1), 67–77. doi:10.1006/jmbi.1999.3244
- Mittl, P. R., Deillon, C., Sargent, D., Liu, N., Klauser, S., Thomas, R. M., Gutte, B., et al. (2000). The retro-GCN4 leucine zipper sequence forms a stable three-dimensional structure. *Proceedings of the National Academy of Sciences of the United States of America*, 97(6), 2562–2566.

- Müller, J., Oehler, S., & Müller-Hill, B. (1996). Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *Journal of molecular biology*, 257(1), 21–29. doi:10.1006/jmbi.1996.0143
- Nair, S. K., & Burley, S. K. (2003). X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell*, 112(2), 193–205.
- Noom, M. C., Navarre, W. W., Oshima, T., Wuite, G. J. L., & Dame, R. T. (2007). H-NS promotes looped domain formation in the bacterial chromosome. *Current biology : CB*, 17(21), R913–4. doi:10.1016/j.cub.2007.09.005
- O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science (New York, N.Y.)*, 254(5031), 539–544.
- O'Shea, E. K., Lumb, K. J., & Kim, P. S. (1993). Peptide “Velcro”: design of a heterodimeric coiled coil. *Current biology : CB*, 3(10), 658–667.
- O'Shea, E. K., Rutkowski, R., & Kim, P. S. (1992). Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell*, 68(4), 699–708.
- O'Shea, E. K., Rutkowski, R., Stafford, W. F., & Kim, P. S. (1989). Preferential heterodimer formation by isolated leucine zippers from fos and jun. *Science (New York, N.Y.)*, 245(4918), 646–648.
- Oehler, S., Eismann, E. R., Krämer, H., & Müller-Hill, B. (1990). The three operators of the lac operon cooperate in repression. *The EMBO journal*, 9(4), 973–979.
- Plesums, J., & Bunch, W. H. (1971). Measurement of phosphorus following <sup>32</sup>P Cerenkov counting. *Analytical biochemistry*, 42(2), 360–362.
- Roy, S., Dimitriadis, E. K., Kar, S., Geanacopoulos, M., Lewis, M. S., & Adhya, S. (2005). Gal repressor-operator-HU ternary complex: pathway of repressosome formation. *Biochemistry*, 44(14), 5373–5380. doi:10.1021/bi047720t
- Rutkauskas, D., Zhan, H., Matthews, K. S., Pavone, F. S., & Vanzi, F. (2009). Tetramer opening in LacI-mediated DNA looping. *Proceedings of the National Academy of Sciences of the United States of America*, 106(39), 16627–16632. doi:10.1073/pnas.0904617106
- Rybenkov, V. V., Vologodskii, A. V., & Cozzarelli, N. R. (1997). The effect of ionic conditions on DNA helical repeat, effective diameter and free energy of supercoiling. *Nucleic acids research*, 25(7), 1412–1418.
- Shore, D., & Baldwin, R. L. (1983a). Energetics of DNA twisting. I. Relation between twist and cyclization probability. *Journal of molecular biology*, 170(4), 957–981.
- Shore, D., & Baldwin, R. L. (1983b). Energetics of DNA twisting. II. Topoisomer analysis. *Journal of molecular biology*, 170(4), 983–1007.

- Shore, D., Langowski, J., & Baldwin, R. L. (1981). DNA flexibility studied by covalent closure of short fragments into circles. *Proceedings of the National Academy of Sciences of the United States of America*, 78(8), 4833–4837.
- Steinmetz, M. O., Jelesarov, I., Matousek, W. M., Honnappa, S., Jahnke, W., Missimer, J. H., Frank, S., et al. (2007). Molecular basis of coiled-coil formation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17), 7062–7067. doi:10.1073/pnas.0700321104
- Thanbichler, M., & Shapiro, L. (2006). Chromosome organization and segregation in bacteria. *Journal of structural biology*, 156(2), 292–303. doi:10.1016/j.jsb.2006.05.007
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular cell*, 10(6), 1453–1465.
- Vinograd, J., Lebowitz, J., Radloff, R., Watson, R., & Laipis, P. (1965). The twisted circular form of polyoma viral DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 53(5), 1104–1111.
- Wang, J. C. (1971). Interaction between DNA and an Escherichia coli protein omega. *Journal of molecular biology*, 55(3), 523–533.
- Watson, J. D., & Crick, F. H. (1953a). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737–738.
- Watson, J. D., & Crick, F. H. (1953b). Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171(4361), 964–967.
- Wiggins, P. A., van der Heijden, T., Moreno-Herrero, F., Spakowitz, A., Phillips, R., Widom, J., Dekker, C., et al. (2006). High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature nanotechnology*, 1(2), 137–141. doi:10.1038/nnano.2006.63
- Wolgemuth, C. W., & Sun, S. X. (2006). Elasticity of alpha-helical coiled coils. *Physical review letters*, 97(24), 248101.
- Yun, K., So, J.-S., Jash, A., & Im, S.-H. (2009). Lymphoid enhancer binding factor 1 regulates transcription through gene looping. *Journal of immunology (Baltimore, Md. : 1950)*, 183(8), 5129–5137. doi:10.4049/jimmunol.0802744